A Novel Aldolase from the Hyperthermophilic Archaeon Sulfolobus solfataricus

Kydd, Catriona L.

Award date:
1999

Awarding institution:
University of Bath

A Novel Aldolase from the Hyperthermophilic Archaeon *Sulfolobus solfataricus*

submitted by

Catriona L. Kydd

for the degree of PhD of the

University of Bath

1999

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGEMENTS

I would like to thank both of my supervisors in the Centre for Extremophile Research, Professor Michael Danson and Dr David Hough, for their continuing support and guidance throughout my PhD. My appreciation also goes to Dr. Christopher Reeve at Avecia Ltd. (previously Zeneca Life Science Molecules) for his enthusiasm and help, especially with the biotransformation work.

Much gratitude is due to both the BBSRC and Zeneca Life Science Molecules for providing funding for what has been an interesting and rewarding project, and for the extra grant money – also very rewarding!

A huge thank you goes to all the members of lab 1.33 for providing motivation to keep going, plenty of entertainment and my royal appointment!

My parents also deserve my thanks, for their help, encouragement and love, which I have always had with me in whatever I have done.

Last, but not at all the least, I would like to thank Andrew, who since my first day here has helped me celebrate the ups and has kept me going through the downs. I hope that as his wife I will always be as supportive.

SUMMARY

Sulfolobus solfataricus is a thermoacidophilic Archaeon that grows optimally at 80°C and pH 3.7. The organism metabolises glucose by way of a modified Entner-Doudoroff pathway that has a net ATP yield of zero. Glucose is dehydrogenated to gluconate and then dehydrated to give 2-keto-3-deoxygluconate (KDG). KDG-aldolase catalyses the reversible aldol cleavage of KDG to give pyruvate and glyceraldehyde. Glyceraldehyde is then further converted to yield a second equivalent of pyruvate.

KDG-aldolase from *S. solfataricus* has been purified and its N-terminal sequence determined. Using this sequence, a DNA probe was obtained by PCR of a genomic DNA library of *S. solfataricus*, and the KDG-aldolase gene cloned and sequenced.

The KDG-aldolase gene encodes a polypeptide of 293 amino acids with a monomeric M_r of 33,000. Sequence identity indicates that the KDG-aldolase enzyme is a member of the NAL superfamily of enzymes that comprises *N*-acetylneuraminate lyase, dihydrodipicolinate synthase, 5-dehydro-4-deoxyglucarate dehydratase and the *Rhizobium meliloti* MosA protein.

The KDG-aldolase gene was expressed in *E. coli* and the recombinant enzyme purified for characterisation studies. KDG-aldolase has been demonstrated to utilise D-, L- and DL-glyceraldehyde with pyruvate in the direction of aldol condensation to produce KDG; the reaction has also been shown to proceed in the *in vivo* direction of aldol cleavage. Thermostability data demonstrate this enzyme to be extremely thermostable, with half-lives of approximately 6 h and 3 h at 95°C and 100°C, respectively. The optimum temperature of activity is 93°C. KDG-aldolase is a class I aldolase, the reaction proceeding via Schiff-base formation between the keto group of the substrate and a lysine residue in the active site of the enzyme. Lys155 of KDG-aldolase aligns with the Schiff-base forming lysine residues of NAL and DHDPS. Similarly to NAL and DHDPS, KDG-aldolase has been demonstrated to exist as a homotetramer. Substrate specificity studies indicate that KDG-aldolase is specific for the

pyruvate donor but variable for the aldehyde acceptor. Crystallisation studies with KDG-aldolase have commenced, with data collected for an orthorhombic crystal to 2.16 Å.

A large-scale biotransformation reaction with KDG-aldolase, producing 5.92 g KDG, representing a 76 % yield, has been performed. The reaction involves the stereospecific carbon-carbon bond formation between pyruvate and DL-glyceraldehyde, forming a new chiral centre and resulting in diastereomeric L-*erythro* and D-*threo* KDG that are separable by anion exchange chromatography.

The characterisation of KDG-aldolase provides valuable information concerning the mechanism and parameters of the enzyme reaction, and will aid further understanding of glucose metabolism in *S. solfataricus*. It is believed that this enzyme has considerable potential for the large-scale production of KDG and other carbohydrates.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ONE LETTER CODE	THREE LETTER CODE	AMINO ACID	ONE LETTER CODE	THREE LETTER CODE	AMINO ACID
A	Ala	Alanine	M	Met	Methionine
C	Cys	Cysteine	N	Asn	Asparagine
D	Asp	Aspartate	P	Pro	Proline
E	Glu	Glutamate	Q	Gln	Glutamine
F	Phe	Phenylalanine	R	Arg	Arginine
G	Gly	Glycine	S	Ser	Serine
H	His	Histidine	T	Thr	Threonine
I	Ile	Isoleucine	V	Val	Valine
K	Lys	Lysine	W	Trp	Tryptophan
L	Leu	Leucine	Y	Tyr	Tyrosine

ATP	adenosine triphosphate
bp	base pair(s)
BSA	bovine serum albumin
Da	daltons
DHDPS	dihydrodipicolinate synthase
DNA	deoxyribonucleic acid
dNTP	deoxy nucleotide triphosphate
EDTA	(disodium) ethylenediamine tetraacetate
FPLC	fast protein liquid chromatography
gDNA	genomic DNA
HPLC	high performance liquid chromatography
IPTG	isopropyl β-D-thiogalactopyranoside
kb	kilobase pairs
kDa	kilodaltons
KDG	2-keto-3-deoxy gluconate
KDGD	2-keto-3-deoxy glucarate dehydratase
LB	Luria-Bertani
M_r	relative molecular mass
NAD(P)	nicotinamide adenine dinucleotide phosphate
NAL	*N*-acetylneuraminate lyase

PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
PCR	polymerase chain reaction
PVDF	polyvinylidene fluoride
RI-HPLC	refractive index high performance liquid chromatography
SDS	sodium dodecyl sulphate
TEMED	N, N, N', N" tetramethylethylene diamine
Tris	tris-(hydroxymethyl)-methylamine
(v/v)	volume / volume
(w/v)	weight / volume (g / 100 ml)
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

".....whatever you do,

do all to the glory of God".

The Holy Bible, New King James Version.

1 INTRODUCTION

1.1 ENZYMES IN BIOCATALYSIS

As the gap between chemistry, biology and engineering narrows, and biotechnology advances, enzyme-catalysed reactions and fermentations are beginning to make appearances in syntheses once dominated by classical chemistry. The ability of an enzyme to perform a task in one step has an immediate advantage over the equivalent chemical reaction that may require many more steps, examples being the formation of a carbon-carbon bond, or the functionalisation of non-activated carbon atoms. Possibly the area in which enzymes are most significantly gaining respect is in their ability both to introduce centres of chirality into optically inactive substrates and to carry out optical resolutions of racemic mixtures. It is becoming increasingly apparent that for every reaction known in organic chemistry, there is a biocatalytic alternative. Table 1.1 gives some examples of the range of large-scale biocatalytic processes.

PROCESS	CATALYST	PRODUCT	ANNUAL TONNAGE
Hydrolases	Amyloglucosidase	Glucose	$10 - 20 \times 10^6$
	Nitrile hydratase	Acrylamide	30×10^3
	Penicillin amidohydrolase	6-aminopenicillanic acid	10×10^3
Resolution (with racemisation)	Hydantoinase *Psuedomonas* sp.	4-hydroxyphenylglycine	100 – 500
		Cysteine	500
Resolution (without racemisation)	Dehalogenase	(*S*)-2-Chloroproionate	2000
Oxidation	Sorbitol dehydrogenase	L-Sorbose	50×10^3
Reduction	β-Ketoreductase	Carnitine	12×10^3
Isomerisation	Glucose (xyl) isomerase	Isoglucose	8×10^6
C-C bond synthesis	Pyruvate decarboxylase	Phenylacetylcarbinol	300 – 500
	Tyrosine phenol lyase	L-Dopa	200
Achiral precursors	Fumarase	Malate	500
	Aspartate ammonia lyase	Aspartate	400
Peptide synthesis	Thermolysin	Aspartame	2000
	Trypsin	Insulin	<1
Glycosyl transfer	Cyclodextrin glucanotransferase	β-Cyclodextrin	800 – 1500

Table 1.1 Examples of large-scale biocatalytic processes. Table reproduced from Turner, (1995).

1.1.1 Advantages and disadvantages of biocatalysts

If biocatalysts are going to make an impact on the industrial preparation of chemicals and other materials then they must have significant advantages over the current method of synthesis and few comparative disadvantages. If the enzyme can do something that the chemical process cannot, for instance asymmetric synthesis of a particular compound that it is imperative to have optically pure, then that is a clear advantage. A large potential market for enzyme catalysts is the pharmaceutical industry, where both asymmetric synthesis and enzymatic chiral resolution of intermediates are of benefit. An example is the biocatalytic resolution of racemic lactam to (-)lactam using an acyclase enzyme, by the company ChiroTech (Cambridge, UK) (McCoy, 1999). Glaxo Wellcome has invested over $40 million

into orders of (-)lactam, which is a key intermediate in the synthesis of abacavir sulphate (Ziagen), an HIV treatment.

Another advantage of some biocatalysts is the potential to increase the rate of a reaction, thus raising the economic value of the process by boosting output and lowering operation expenses. This is the case with flavin reductase from *Rhodococcus*, an enzyme that activates the biodesulphurisation pathway, increasing the rate ten-fold. Energy BioSystems (The Woodlands, Texas, USA), have patented this application for increasing the yield of oil molecules from the oil base in the petroleum industry (D'Aquino, 1999).

Decreasing the environmental impact can strengthen the case for biocatalysts, as they are generally cleaner than their metal catalyst counterparts. The amount of waste product generated can be reduced, as is the case with the conversion of adiponitrile to 5-cyanovaleramide, a herbicide, by A4-nitrohydroamylase. Nitto Chemical Industries (Tokyo, Japan) have eliminated the use of manganese dioxide, thus reducing the amount of waste per product formed by 99.5 % (w/w) (D'Aquino, 1999). Another advance, although not strictly an enzyme catalysed one, is in the biopolymer industry and the replacement of petroleum-based feedstocks with materials from renewable resources that are largely biodegradable. Polyhydroxybutyrate biopolymers (PHB) are made by fermentation of recombinant *Alcaligenes eutrophus* growing on glucose and propionic acid (Kim et al., 1994). Monsanto (St. Louis, Missouri, USA) uses BIOPOL™ (containing PHB) for the manufacture of credit cards issued by Greenpeace.

A further benefit of investigation into enzyme technologies is the enormous range of activities and characteristics available in nature. Screening for the desired reaction from expression libraries prepared from environmental DNA may deliver an enzyme perfectly suited to particular industrial requirements. Alternatively, a natural advantage of biocatalysts over chemical syntheses, is the ability to adapt an enzyme to the process required. Various directed evolutionary techniques, such as random mutagenesis and DNA shuffling, have

yielded enzymes with 'improved' characteristics desirable for the reaction conditions. These techniques will be discussed further in section 1.1.5.

The economics of a process defines whether or not it will be taken up by industry, regardless of how attractive the advantages. Presently, the most significant disadvantage of biocatalysts is the capital required to cross the entry barrier of high set-up costs; for instance, the relative cost of enzymes compared to metal catalysts is frequently high. Other disadvantages include the inherent instability of most enzymes, product inhibition and low rates of catalysis. However, each of these can be overcome by selecting for alternative or novel enzymes with more favourable products, or by mutagenesis of existing enzymes.

Table 1.2 summarises the requirements to be examined before a biocatalyst is chosen to replace a chemical synthesis.

Discovery scientists	Novel science
⇩	
Development	Reliable, reproducible and easy process integration
⇩	
Production	Minimal commissioning problems, speed to first production
⇩	
Company	Good returns on investments, low capital costs
⇩	
Consumer	New or better quality products and services at more affordable prices
⇩	
Society	Job creation, export creation
⇩	
The planet	Use of renewable resources, environmental benefits

Table 1.2 Requirements to produce a successful biocatalyst. Reproduced from Cheetham (1998).

1.1.2 Extremozymes

1.1.2.1 Extremophiles and their extreme environments

Whilst some extremozymes (enzymes able to withstand extreme conditions) have been found in organisms that have growth optima characteristic of mesophiles, more commonly these enzymes are derived from organisms inhabiting environments once believed to be hostile to all forms of life. With the improvement of culture conditions however, microorganisms designated 'extremophiles' were found to be thriving within such sites as hot springs, cold Arctic water, acidic and alkaline water, saturated salt brines and deep-sea hydrothermal vents. According to each extremophile's optimal growth conditions they are named thermophiles, psychrophiles, acidophiles, alkaliphiles, halophiles and barophiles, respectively (Table 1.3).

PHENOTYPE	ENVIRONMENT	TYPICAL GENERA
Thermophilic	55 – 80 °C	*Methanobacterium, Thermoplasma, Thermus**, some *Bacillus** species
Hyperthermophilic	80 – 113 °C	*Aquifex**, *Archaeoglobus*, *Hydrogenobacter**, *Methanothermus*, *Pyrococcus*, *Pyrodictium*, *Pyrolobus*, *Sulfolobus*, *Thermococcus*, *Thermoproteus*, *Thermotoga**
Psychrophilic	-2 – 20 °C	*Alteromonas**, *Psychrobacter**
Halophilic	2 – 5 M NaCl	*Haloarcula*, *Halobacterium*, *Haloferax*, *Halorubrum*
Acidophilic	pH < 4	*Acidianus*, *Desulfurolobus*, *Sulfolobus*, *Thiobacillus**
Alkaliphilic	pH > 9	*Natronobacterium*, *Natronococcus*, some *Bacillus** species

Table 1.3 Extremophiles and their environments. Genera indicated with an asterisk (*) are of the domain Bacteria; all others are Archaea. Reproduced from Hough and Danson (1999).

It is interesting that some extremophiles display combinations of phenotypic traits, which may be linked in terms of adaptation of cellular components. For example, thermoacidophiles prefer not only acid pH values but also elevated temperatures. Similarly,

barophiles proliferate in highly pressurised abyssal waters that are at temperatures near 0°C (Adams et al., 1995).

1.1.2.2 The Archaea

Prior to 1977, definition of taxonomy had existed as a phenotypic classification giving either a five-kingdom organisation or the eucaryotic/procaryotic dichotomy. Phylogenetic analysis of ribosomal RNA sequences by Woese and Fox (1977) revealed a molecular level system of taxonomy that gave a tripartite view of evolutionary lineage. This system was then further refined to give the three domains known as the Bacteria, the Archaea and the Eucarya (Woese and Olsen, 1986; Woese et al., 1990) (Figure 1.1). The archaeal domain can be further subdivided into the kingdoms euryarchaeota (the methanogens and their phenotypically diverse relatives) and crenarchaeota (the extreme thermophiles) (Woese et al., 1990).

Whilst there are certainly extremophilic organisms of bacterial origin, the vast majority of organisms with these interesting phenotypes are members of the archaeal domain. However, it is also important to be aware that not all Archaea are extremophiles (Jurgens and Saano, 1999).

1.1.3 Extremozymes and their potential as biocatalysts

With the advent in the last fifteen years of biochemical knowledge of enzymes from extremophilic organisms, comes the realisation of the biotechnological advantage that these enzymes may have over their conventional counterparts from mesophiles. These hyperstable enzymes are able to withstand extremes of temperature, pH, salinity, pressure and solvent conditions. Table 1.4 details the range of conditions inhabited by the organisms that these enzymes come from, and gives examples of their industrial applications.

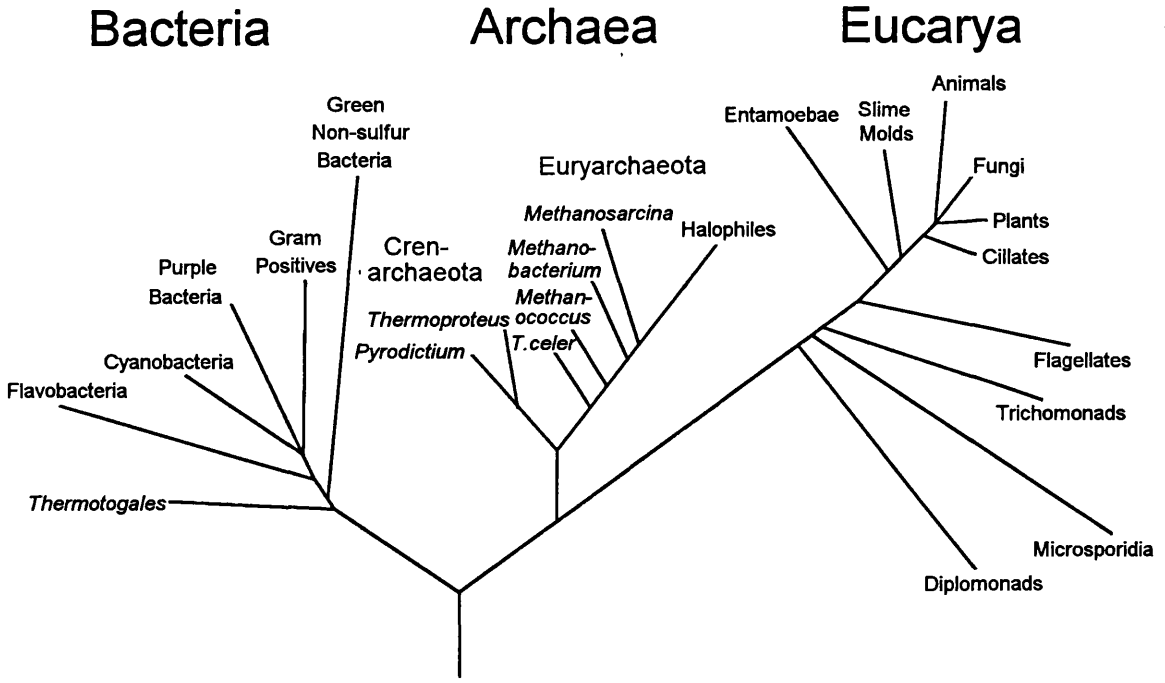


Figure 1.1 The three domain universal phylogenetic tree as proposed by Woese et al. (1990). Figure reproduced from Wheelis et al. (1992).

ORGANISM	ENVIRONMENT	INDUSTRIAL APPLICATIONS OF RELATED ENZYMES
Psychrophile	-2 to 20°C	Food processing, fragrances, laundry detergents
Thermophile	55 to 113°C	Corn syrup, oil recovery, drug development
Halophile	2 to 5 M NaCl	Chiral purification, pharmaceuticals, oil recovery
Acidophile	pH < 4	Synthesis of compounds in acidic solutions; animal feed
Alkaliphile	pH > 9	Detergents, textiles ("stone-washed" jeans)

Table 1.4 Some examples of industrial applications of enzymes from extremophilic organisms. Reproduced from Diversa, Inc. (San Diego, CA, USA; internet page http://www.biocat.com/).

1.1.3.1 Structural basis of enzyme stability

The ability of extremozymes to withstand such extreme conditions makes them attractive candidates for biotechnological exploitation beyond the traditional uses of enzymes in the starch processing and detergent industries. It is interesting to understand therefore what imparts this stability to the protein, and if possible, how it can be engineered or mimicked to the advantage of the biotechnology industry. This has been a popular area of investigation, and it has become apparent that there is no definite pattern or recipe for creating more stable proteins. Instead, changes in the stability of extremozymes are due to a variety of subtle interactions that combine to give global changes in structure. Examples of these individual changes, with regard to thermostability, are given below:

- Increased resilience and/or rigidity - Important for protecting against unfolding and preserving the catalytic nature of the enzyme. Resilience is conferred to by polar interactions, hydrogen bonds and ion-pairs; rigidity is conferred to by packing efficiency, secondary structure features and amino acid substitutions (Aguilar et al., 1997).

- Increased hydrophobic interactions - Hydrophobic interactions contribute to the decrease in Gibb's free energy required for the protein to fold and remain folded in aqueous solutions (Pace, 1992; Kirino, 1994).
- Increased packing efficiency - This results in the reduction of the surface area : volume ratio (Chan et al., 1995) and/or a decrease in the volume of internal cavities (Russell et al., 1997).
- Increase in ionic interactions - Both inter-domain and intra-subunit ion pairs and ion pair networks contribute to stability (Perutz and Raidt, 1975; Russell et al., 1997).
- Secondary structure changes - Shorter loop regions increase the compactness and reduce the flexibility of the enzyme, thereby increasing the overall rigidity. Inclusion of stabilising factors such as Ala in helices and Pro in shortened loops also contributes to overall stability (Matthews et al., 1987; Menéndez-Arias and Argos, 1989; Russell et al., 1997).
- Resistance to covalent destruction - Fewer residues susceptible to deamidation (Asn, Gln) or oxidation (Cys) also characterise an increase in thermostability (Hensel et al., 1992; Russell et al., 1997).
- Increase in hydrogen bonds - An increase in the number of intra-subunit neutral-neutral and especially charged-neutral hydrogen bonds correlates to an increase in thermostability by contributing to the overall resilience (Tanner et al., 1996).

It is evident that for each protein that employs one combination of mechanisms for increasing its thermal stability there will be another protein that uses an entirely different combination of factors.

In addition to enzymes from thermophiles, examination of enzymes from other extremophiles indicates a similar pattern for conferring stability. Cold-active enzymes have an increased flexibility, extended surface loop regions, and reduced hydrophobic interactions

compared to their mesophilic counterparts. An increase in intra-molecular ion-pairs, even compared to a hyperthermophilic homologue, may be to prevent cold-induced unfolding (Russell et al., 1998).

Halophilic enzymes are characterised by a high degree of negatively charged amino acids, with an excess of these acidic residues at their surfaces. This enables formation of a hydrated ion network over the protein surface, thereby reducing surface hydrophobicity and the tendency to aggregate. There may also be an increase in the number of ion pairs and the introduction of stabilising residues in α-helices (Dym et al., 1995).

There have been no acid or alkali stable enzymes that have been structurally characterised to date and therefore it is not possible to define criteria for their stabilisation. It is important to note that in the case of acidophiles or alkaliphiles, the intracellular environment is close to neutral and so intracellular enzymes may be relatively unchanged compared to similar enzymes from mesophiles.

1.1.4 Examples of the use of biocatalysts in industry

1.1.4.1 Bioremediation

An exciting advance in the area of bioremediation is the degradation of organopollutants in radioactive mixed waste environments. Solely physicochemical methods are not suitable for the clean-up of these wastes due to their extreme toxicity and the high cost involved, so an alternative method must be found. *Deinococcus radiodurans* is able to survive high-level exposures to gamma irradiation inducing more than 150 double-stranded breaks per chromosome due to a remarkably efficient *recA*-dependent repair system (Daly et al., 1994). *D. radiodurans*, expressing recombinant toluene dioxygenase genes from *Pseudomonas putida* F1, has been demonstrated to oxidise toluene, chlorobenzene, 3,4-dichloro-1-butene and indole, whilst withstanding a highly irradiating environment and the solvent effects of toluene and thrichloroethene (Lange et al., 1998). Although not presently employed for the bioremediation of organopollutants from radioactive soil and water, the

experiments by Lange et al. (1998) do point the way to a solution to this problem, and suggest the potential for cloning further genes encoding degradative enzymes into *D. radiodurans*.

1.1.4.2 Textile industry

An interesting example of harnessing the ability of enzymes in the textile industry is in the production of indigo for dyeing blue jeans by Genencor International (GCI, South San Francisco, CA, USA) (Ensley et al., 1983). The key enzyme used is napthalene dioxygenase (NDO), an enzyme in the tryptophan biosynthetic pathway of *Pseudomonas putida* that converts indole to indoxyl, which then undergoes oxidative dimerisation to form indigo. The gene for napthalene dioxygenase was cloned into *E. coli*; however, it was also necessary to genetically modify the organism to ensure that the extra glucose feedstock was committed to indigo production, and that the indigo produced was not further modified. A final production strain of *E. coli* containing over 18 kb of additional, co-ordinately regulated DNA in 15 open reading frames was developed. The method of producing recombinant indigo uses less than half the unit operations compared to synthetic indigo, which uses and produces highly toxic chemicals. Moreover, recombinant indigo can be produced at a cost equal to the synthetic dye, making Genencor's method far more appealing.

In 1996 the worldwide market for laundry detergents was just under $500 million, accounting for a third of the total market for industrial enzymes (http://www.dyadic-group.com/enzymes.htm). The enzymes most commonly included in detergents are proteinases and amylases for the degradation by hydrolysis of protein and starch residues, such as blood, sweat or food products, on the clothes. More recently, enzymes such as lipases are added, to improve the removal of fatty food stains and sebum at lower temperatures from clothes. Cellulases are also added for softening and colour-brightening purposes (Christensen et al., 1987). The use of enzymes from psychrophilic organisms opens the door on a new range of low-temperature, energy (and cost) efficient laundry-detergents.

1.1.4.3 Pharmaceutical industry

Antibiotics account for 11 % ($32.5 billion) of the global pharmaceuticals market and is growing at 4 – 5 % / yr (D'Aquino, 1999). The use of new biosynthetic routes in the manufacture of antibiotics and antibiotic precursors stands to improve the economics of the pharmaceutical industry significantly. DSM (Heerlen, Netherlands) is planning the production of 7-amino-deacetoxy-cephalosporanic acid (7-ADCA) via a fermentation route. 7-ADCA is a precursor for cephalosporin, which itself can be modified enzymatically to make other semi-synthetic cephalosporin based antibiotics. DSM explains that the advantages are the fewer reaction steps in the biosynthetic pathway, and the absence of organic solvents required by the conventional chemical route.

Altus Biologics Inc., (Cambridge, Mass, USA) has also developed a preferential route for the biosynthetic manufacture of antibiotics such as ampicillin and amoxicillin. The enzyme penicillin acyclase selectively cleaves the phenylacetic-acid side chain from penicillin-G to produce the key intermediate 6-aminopenicillanic acid. A new amino acid side chain can then be linked to form the novel antibiotic. The chemical synthesis is not catalytic and operates at -40°C in organic solvents. This biosynthetic method has clear environmental and economic advantages as it runs in aqueous solution at neutral pH and ambient temperatures (D'Aquino, 1999).

1.1.4.4 Food processing industry

The two main enzyme uses within the food processing industry are starch degrading enzymes for dextrin and sugar syrup production, and animal/microbial rennets for cheese manufacture. In 1988 these accounted for approximately $100 million and $65 million respectively of a $500 million market. A summary of the range of enzymes used in food processing is given in Table 1.4.

ENZYME	MAJOR SOURCE	APPLICATIONS
Amylase	*Bacillus subtilis* *Aspergillus* sp Barley	Glucose syrup production, brewing, fruit processing, baking
Cellulase	*Aspergillus niger* *Trichoderma reesei*	Fruit processing, flavour production
Dextranase	*Penicillium* sp *Chaetomium* sp	Sugar processing
Glucanase	*Bacillus subtilis* *Penicillium emersonii* *Trichoderma reesei*	Brewing, wheat processing
Lactase (β-galactosidase)	*Aspergillus* sp *Bacillus* sp *Saccharomyces* sp	Milk processing
Lipase	*Aspergillus niger* *Candida* sp *Mucor* sp *Rhizopus* sp	Flavour production, enzyme modified cheese, fat modification, emulsifier synthesis
Pectinase	*Aspergillus* sp	Fruit processing
Protease	Animal e.g. trypsin *Aspergillus* sp *Bacillus* sp Plant e.g. papain	Baking, brewing, protein hydrolysates, flavour production
Xylanase (pentosanase)	*Aspergillus* sp *Trichoderma* sp	Wheat processing, baking

Table 1.5 Uses of enzymes in the food industry. Table reproduced from West (1988).

The manufacture of high-fructose corn syrups (HFCS) is a major application of the market for starch hydrolysing enzymes. HFCS are used to replace sucrose syrups in food and beverages due to their high sweetening properties. The production of HFCS uses four enzymes in three steps as summarised in Table 1.6.

STEP	ENZYME(S)	TEMPERATURE	PH	TIME
Dextrinisation (Liquefaction)	α-Amylase	95°C 105°C	6.5	1 – 2 h 5 – 10 min
Saccharification	Glucoamylase, Pullulanase	60°C	4.5	hours
Isomerisation	Glucose isomerase	60 - 65°C	7.5	hours

Table 1.6 The conditions and enzymes used for production of HFCS.

High operating temperatures are required for the complete solubilisation of starch, to minimise contamination and to decrease the polymerisation of D-glucose to *iso*-maltose (Nigam and Singh, 1995). The enzymes involved in the saccharification, and to some degree dextrinisation (also called liquefaction), must also be able to withstand acid pH. The necessity for heat and acid stable enzymes has instigated a large amount of research into the relevant enzymes from thermophilic host organisms; examples are, pullulanase and glucoamylase from *Clostridium thermohydrosulfuricum,* (Hyun and Zeikus, 1985); α-amylases from *Pyrococcus woesei* (Koch et al., 1991) and *Pyrococcus furiosus* (Dong et al., 1997) and glucose isomerase from *Thermotoga* sp (Brown et al., 1993; Vieille et al., 1995).

1.1.4.5 Petroleum industry

A crucial step in well stimulation for oil and gas recovery is the hydraulic fracturing of the well-bore to improve recovery. Flooding the well-bore with a viscous, water-based fluid containing guar gum, a galactomannan-based polymer, achieves this. This fluid also contains proppants, particles used to hold open crevices generated by the high levels of hydrostatic pressure applied to the flooded well. To enable the oil or gas to flow into the well, the viscosity of the fracturing fluid must be reduced *in situ*. This has been carried out by chemical oxidisation or by enzymatic hydrolysis of the polymer using hemicellulases. Both methods have the disadvantage of prematurely reducing the viscosity of the fluid before

stimulation of the well has been accomplished. The enzymatic method has the additional disadvantage of a severe reduction in activity at the high temperatures, frequently over 120°C, within the well-bore.

The discovery and application of thermoactive and thermostable enzymes have the potential to revolutionise this process. α-Galactosidase and β-mannanase from the hyperthermophilic bacterium *Thermotoga neapolitana*, are able to hydrolyse the galactomannan chain of the guar gum. These enzymes withstand the high temperatures and pressures in the well-bore, but have significantly reduced activities at ambient temperatures, ensuring little hydrolysis during the fluid preparation or addition to the well (McCutchen et al., 1996).

1.1.5 Screening for novel biocatalysts and development of existing ones

As the majority of enzymes are functional within the parameters of the growth conditions of the organism from which they are derived, it makes sense to screen ecological niches characteristic of the qualities desired in the enzyme. For example, if a heat-stable, alkali-tolerant protease is required for a laundry detergent, it may be advantageous to screen environmental samples from Africa soda lakes for a protease producing thermoalkaliphile. However, it is a severe handicap that fermentation conditions mimicking the natural habitat are often toxic (sulphur-reducing hyperthermophiles), corrosive (sulphur-reducing hyperthermophiles, acidophiles, halophiles) or impractical (barophiles). This is reflected by the generally accepted fact that the proportion of microorganisms that have actually been cultured using conventional techniques is less than 1 % of those existing in the biosphere (Amann et al., 1995; Pace, 1997). An alternative to screening enzymes purified from pure cultures of microorganisms, is the screening of expression libraries prepared from environmental DNA by PCR using degenerate primers (Okuta et al., 1998).

A different approach altogether to screening for the enzyme that fits the job description is to take an enzyme that meets some of the requirements and to modify it such

that it develops novel, or improved characteristics. Using the array of criteria that contribute to protein thermal stability, or with knowledge about an enzyme's active site environment, a rational site-directed mutagenesis programme may yield mutants with increased stability, or altered specificity.

Recently developments for obtaining enzymes with altered characteristics desirable for specific biotechnological applications include various evolutionary methods. These techniques are based on the creation of libraries of genes with low levels of mutations by error-prone PCR. Either a selection or screening method is used to identify mutants with improvement of the desired property. The gene(s) encoding these modified enzymes are then subjected to further cycles of error-prone PCR and selection / screening such that beneficial mutations are accumulated (Kuchner and Arnold, 1997). This technique has been used to enhance the activity of subtilisin E in organic solvents (Chen and Arnold, 1993; You and Arnold, 1996).

DNA shuffling is another technique that has been applied as an extension to directed evolution. Sets of homologous genes improved by directed evolution are randomly fragmented and re-assembled in a self-priming polymerase reaction. This allows a more rapid accumulation of beneficial mutations identified in separate genes (Moore et al., 1997).

Alternatively, DNA shuffling using families of genes from diverse species has been used to accelerate the improvement of moxalactamase activity in cephalosporinases. A fifty fold increase in activity per cycle of shuffling was obtained when the four genes were available in a mixed pool, compared to when they were evolved separately (Crameri et al., 1998). Clearly, directed evolution and DNA shuffling have a significant contribution to make to the development of enzymes for industrial biotransformation processes.

1.2 ALDOLASES AND CARBON-CARBON BOND FORMATION IN BIOCATALYSIS

An application of particular interest to the field of biocatalysis is the synthesis of water-soluble, polyfunctional organic molecules, such as carbohydrates and carbohydrate analogues. Conventional synthetic approaches to these compounds are usually complex, involving multiple protection and deprotection steps, and yielding racemic mixtures of products. The use of enzyme catalysts has the potential to provide high degrees of stereoselectivity under mild conditions for the production of complex carbohydrates and their conjugates with various biological and associated functions (Takayama et al., 1997).

Aldolases, originally named zymohexases, are ubiquitous enzymes that have been purified from a large variety of animals, plants and microorganisms (Horecker et al., 1972). These enzymes are so named due to their catalysis of the reversible aldol condensation between a carbonyl-group-containing donor unit, for which they are usually specific, and a variety of acceptor aldehydes (Takayama et al., 1997). An almost entirely common feature of the aldolase condensation reaction is that the enzyme controls the stereochemistry of the C-C bond formation. The fact that the reaction is irrespective of the stereochemistry and structure of the substrates enables the predictable and asymmetric synthesis of various carbohydrates.

1.2.1 Aldolase groups and reaction mechanisms

Aldolase enzymes can be grouped according to the structure of the donor substrate (Figure 1.2). The four groups are dihydroxyacetone phosphate-dependent aldolases, pyruvate- or phosphoenolpyruvate pyruvate-dependent aldolases, acetaldehyde-dependent aldolases, and glycine-dependent aldolases. It is a feature of most aldolases that they are specific for their donor substrate, but display versatility with regard to their acceptor aldehyde (Takayama et al., 1997). However, there are occasional exceptions to this rule, such as is the case with 4-hydroxy-2-keto-pentanoic acid aldolase from *E. coli*, that is highly specific for the acetaldehyde acceptor, but variable for the donor (Pollard et al., 1998).

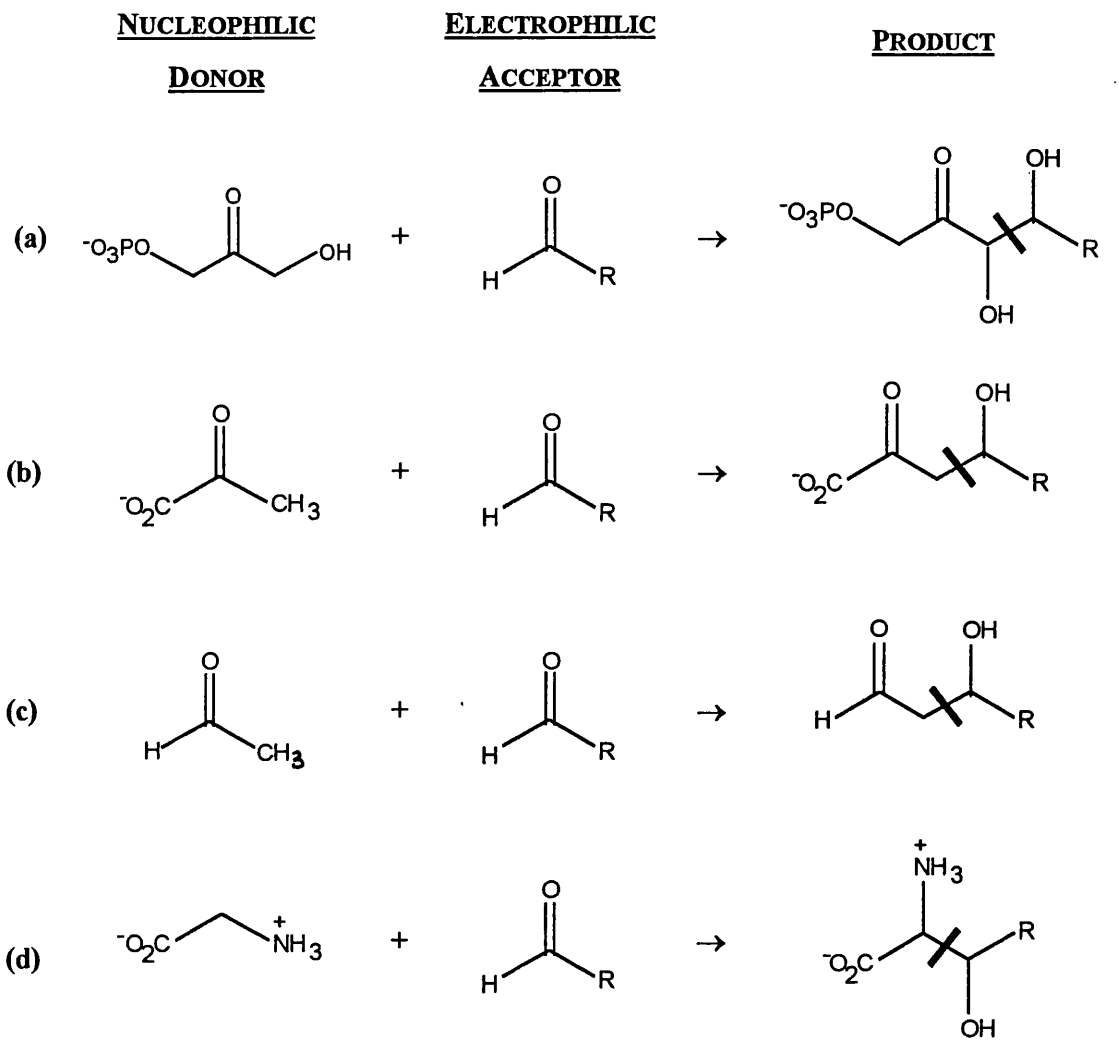


Figure 1.2 Aldolase enzymes grouped according to nucleophile type: (a) dihydroxyacetone phosphate-dependent aldolases, (b) pyruvate- or phosphoenolpyruvate pyruvate-dependent aldolases, (c) acetaldehyde-dependent aldolases, and (d) glycine-dependent aldolases. The thick line across the bond of the product indicates the bond formed (adapted from Wong et al., 1995; Shelton et al., 1996).

The largest group of aldolases is that which uses pyruvate or phosphoenolpyruvate as the nucleophilic donor constituent of the reaction (Shelton et al., 1996). A conserved four-carbon chain, with a 4-hydroxy-2-ketobutyrate framework in the case of pyruvate, is produced with four contiguous and differentially functionalised carbon atoms. This framework provides opportunities for various chemical transformations in aqueous solution without the need for protecting groups, such as the synthesis of α-amino-γ-hydroxycarboxylic acids, β-hydroxycarboxylic acids, α,γ-dihydroxycarboxylic acids, and 2-deoxy-aldose sugars (Figure 1.3), thus making this group of aldolases an attractive target for biosynthesis of optically pure compounds

Aldolases from all four groups can also be subdivided into two classes according to their reaction mechanism (Rutter, 1964). Class I aldolases form a Schiff-base intermediate with the keto-oxygen of the donor substrate, via a lysine residue in the active site (Figure 1.4). They can be inactivated by sodium borohydride in the presence of substrate and are unaffected by EDTA. Class II aldolases, however, use a divalent metal ion cofactor, acting as a Lewis acid, in the active site of the enzyme (Figure 1.5). Sodium borohydride has no effect on these enzymes in the presence or absence of substrates, but they are inhibited by EDTA, indicating that a metal ion cofactor is required. Class I enzymes are predominantly found in animals and higher plants, whereas class II enzymes are found in bacteria and fungi (Rutter, 1964). Aldolases from the Archaea have been found to be of both classes (Dhar and Altekar, 1986; De Montigny and Sygusch, 1996).

1.2.2 Examples of aldolases used in organic synthesis

The dihydroxyacetone phosphate-dependent aldolases are the most extensively studied group of aldolases, with the majority of investigation having been directed towards fructose 1,6-bisphosphate aldolase (Takayama et al., 1997). One application of this group of

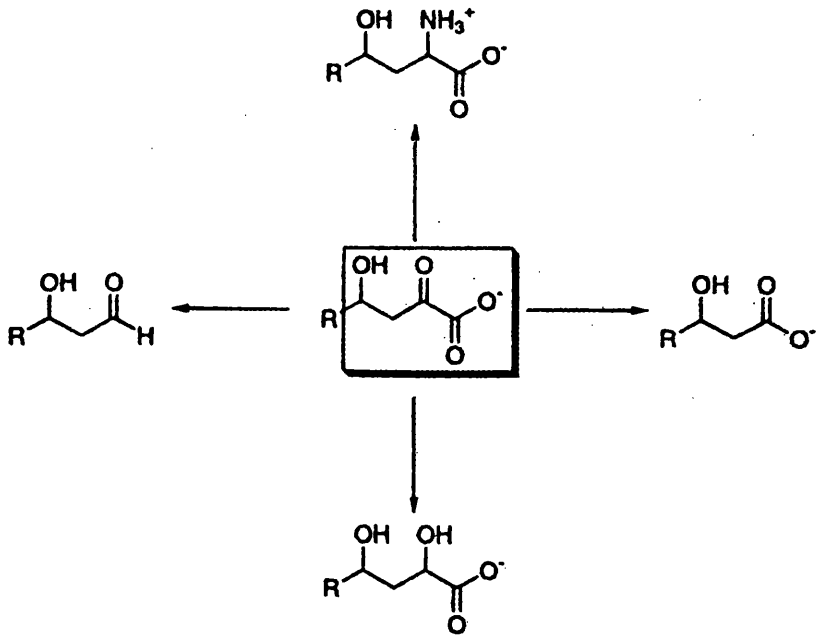


Figure 1.3 Possible transformations using products from pyruvate aldolase condensation reactions. Clockwise form the top; α-amino-γ-hydroxycarboxylic acids, β-hydroxycarboxylic acids, α,γ-dihydroxycarboxylic acids, and 2-deoxy-aldose sugars. Figure reproduced from Shelton et al., 1996.

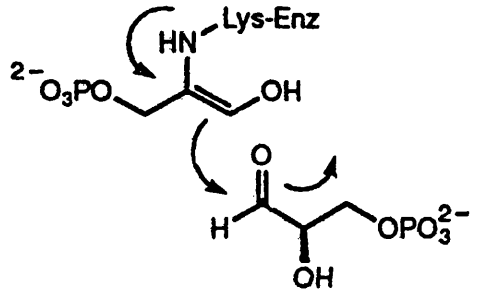


Figure 1.4 Proposed reaction mechanism for a Class I dihydroxyacetone phosphate dependent aldolase. Figure reproduced from Wong et al. (1995). For further mechanistic detail of Class I aldolases refer to Figure 7.5.

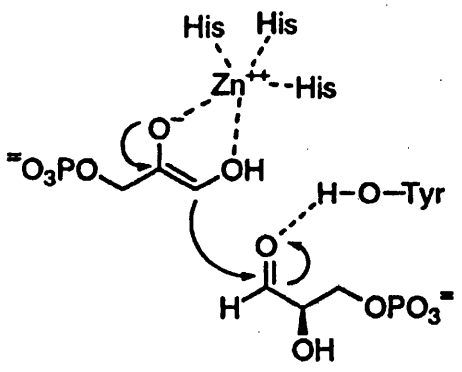


Figure 1.5 Proposed reaction mechanism for a Class II dihydroxyacetone phosphate dependent aldolase. Figure reproduced from Takayama et al. (1987).

aldolases has been the synthesis of azasugars for use as glycosidase and glycosyltransferase inhibitors. Inhibition of these enzymes has implications for both anti-tumour and anti-inflammatory therapies. The synthesis takes advantage of the low specificity aldolases have for their aldehyde acceptor substrates in the stereospecific aldol condensation of dihydroxyacetone phosphate with an azido aldehyde, followed by reductive amination to give various azasugars (Qiao et al., 1996).

N-acetylneuraminate lyase (NAL), also known as *N*-acetylneuraminic acid aldolase or sialic acid aldolase, is one of the best-characterised aldolase enzymes. The enzyme catalyses the reversible condensation of pyruvate with D-*N*-acetylmannosamine (ManNAc) to form *N*-acetylneuraminic acid (Neu5Ac) (Figure 1.6).

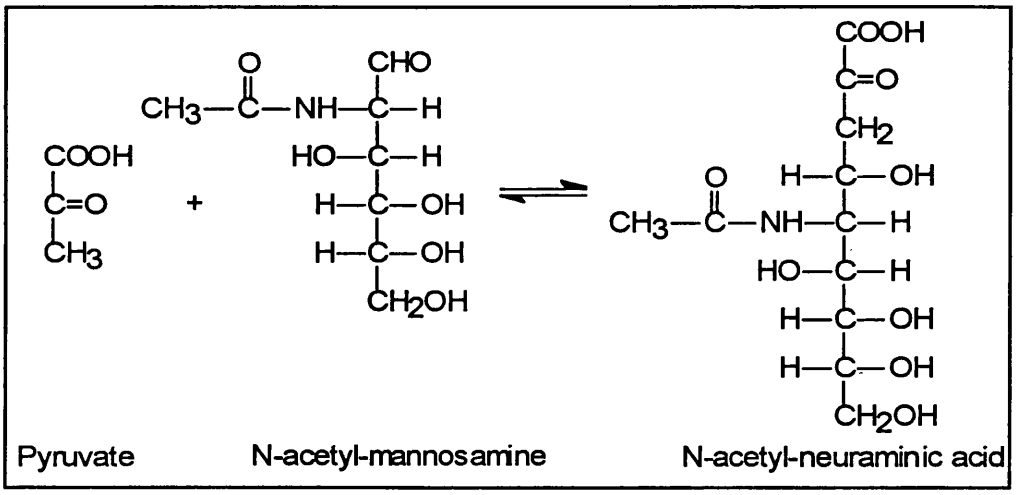


Figure 1.6 ***N*-acetylneuraminate lyase catalysed reaction.**

Neu5Ac and derivatives of Neu5Ac are termed sialic acids and are found at the termini of mammalian glycoconjugates, where they play an important part in biochemical recognition (Corfield and Schauer, 1982). These ligands are recognised by the influenza virus and are bound by the proteins haemagglutinin and sialidase, which are present on the virus surface. Synthesis of Neu5Ac and its derivatives is of interest for the inhibition of sialidase and haemagglutinin and has been investigated using the NAL enzyme (Sparks et al., 1993; Von Itzstein, 1993).

1.3 KDG-ALDOLASE FROM SULFOLOBUS SOLFATARICUS

1.3.1 Sulfolobus solfataricus

S. solfataricus (formerly named *Caldariella acidophila*), a member of the Crenarchaeota order of the Archaeal domain, was initially isolated from hot, acidic pools in the Pisciarelli solfatara crater north of Naples, Italy (De Rosa et al., 1975). The thermal water had a temperature range of 74 to 89°C and a pH range of 1.4 to 2.6 due to an abundance of sulphuric acid in the pool, and sulphur on the water surfaces and in the surrounding soil. The strain isolated, MT4, had a growth range of 63 to 89°C, a growth optimum of 87°C, and did not grow at 60 or at 92°C. At 87°C a broad pH optimum was measured from pH 3.0 to pH 4.5. No growth was observed at pH 1.0 or pH 5.5 (De Rosa et al., 1975). A more recent isolate, P1 (DSM 1616), has a similar temperature range of growth, but is optimal at 80°C. The pH range was more defined over the range of pH 2.0 to pH 4.5, with an optimum at pH 3.7 (Grogan, 1989). *S. solfataricus* is a strictly aerobic facultative autotroph, able to grow hetereotrophically on yeast extract or autotrophically on elemental sulphur. The cell morphology is generally presented as irregular lobed spheres. The GC % content is characteristically low for all Sulfolobales, and is about 37% for *S. solfataricus* (Segerer and Stetter, 1991).

1.3.2 Entner-Doudoroff pathway

S. solfataricus is understood to metabolise glucose via a modified version of the Entner-Doudoroff pathway, producing two equivalents of pyruvate with a net ATP yield of zero (Entner and Doudoroff, 1952; De Rosa et al., 1984; Selig et al., 1997). The classical Entner-Doudoroff pathway is common to aerobic Bacteria, with a modified version involving non-phosphorylated substrates being more prevalent amongst aerobic Archaea (Kersters and De Ley, 1968; Selig et al., 1997). The modified Enter-Doudoroff pathway of *S. solfataricus* is illustrated in Figure 1.7.

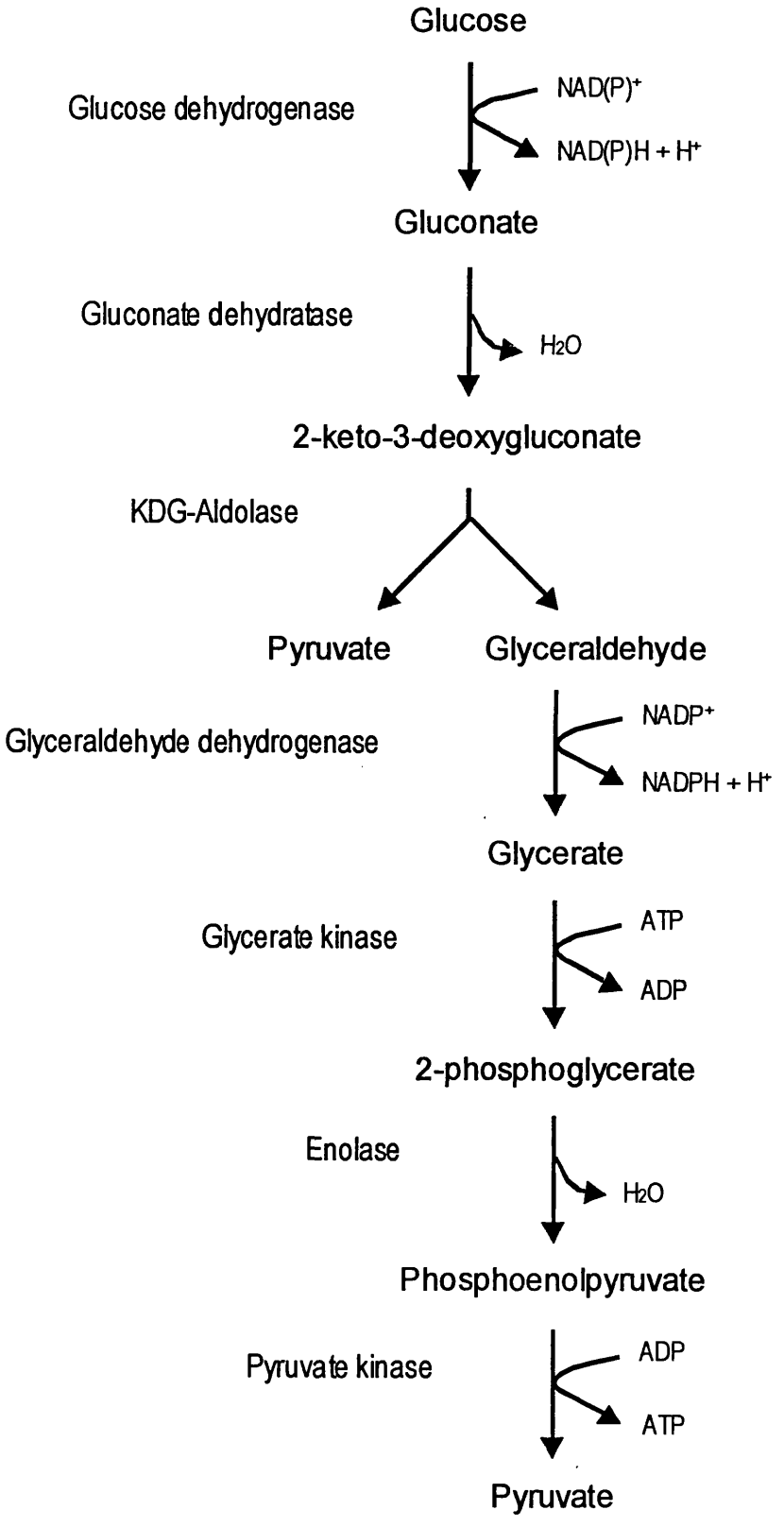


Figure 1.7 The modified Entner-Doudoroff pathway of *Sulfolobus solfataricus*.

Glucose is converted to gluconate by an $NAD(P)^+$ dependent glucose dehydrogenase, and is then dehydrated by gluconate dehydratase to give 2-keto-3-deoxygluconate (KDG). KDG is cleaved by KDG-aldolase, yielding the first equivalent of pyruvate plus glyceraldehyde. Glyceraldehyde is then believed to be converted to glycerate by an $NAD(P)^+$ dependent dehydrogenase and then phosphorylated to give 2-phosphoglycerate, which is further converted to give a second equivalent of pyruvate via enolase and pyruvate kinase. One mole of ATP is consumed in the phosphoglycerate kinase reaction, and one mole is generated in the pyruvate kinase reaction, giving a net yield of zero (De Rosa et al., 1984; Selig et al., 1997). This same modified pathway has also been identified in the thermoacidophilic Archaeon *Thermoplasma acidophilum* (Budgen and Danson, 1986).

1.3.3 KDG-aldolase catalysed conversion of KDG to pyruvate and glyceraldehyde

The reaction of KDG to pyruvate and glyceraldehyde in *S. solfataricus* was identified by De Rosa et al. (1984), by enzymatic detection of the products pyruvate and glyceraldehyde (Figure 1.8). The reaction was not found to be affected by Mg^{2+} or Mn^{2+} ions, or by ATP.

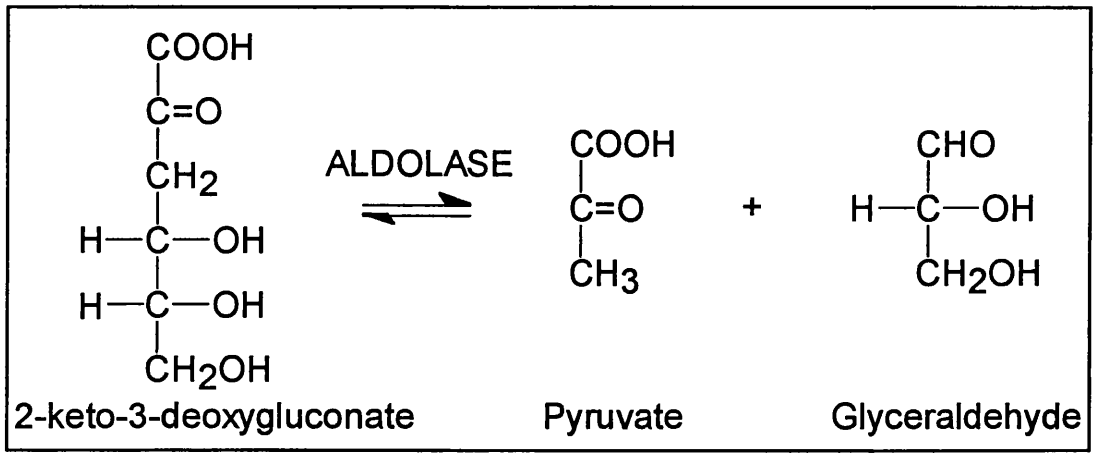


Figure 1.8 KDG-aldolase catalysed reaction.

KDG-aldolase has a catabolic function *in vivo* and the reaction occurs in the retro-aldol direction. However, the reaction is reversible and can also be used to produce KDG from the substrates pyruvate and glyceraldehyde. Augé and Delest (1993) give an equilibrium

constant of 0.33 mM for the *in vitro* reaction with KDG-aldolase from the filamentous fungus *Aspergillus niger*, indicating that at equilibrium there would be 27 % of each of pyruvate and glyceraldehyde, and 46 % of KDG.

1.3.4 Potential applications of KDG-aldolase

KDG is a compound that is not at present commercially available. There are numerous references in the literature for it use in various small-scale investigations, such as for the elucidation of the modified Entner-Doudoroff pathway in *S. solfataricus* (De Rosa et al., 1984). Each time that a group has required KDG it has been necessary to synthesise and purify the compound, a process that is often complicated (see chapter 9). A large-scale production using KDG-aldolase has the potential to provide sufficient KDG for commercial exploitation using readily available substrates.

KDG-aldolase has also been identified in the filamentous fungus *Aspergillus niger* when grown on glucose as the sole carbon source (Augé and Delest, 1993). Resting mycelium cells have been used for the stereospecific production of KDG, although the purity of the product was compromised by additional enzyme activity in the whole cell reaction. A biotransformation with purified KDG-aldolase would be preferential for the production of KDG.

KDG, like other products of the pyruvate aldolases (Figure 1.2), is densely and differentially functionalised, having five carbon atoms with different oxidation states. A variety of potentially interesting compounds may be synthesised using KDG as the starting material. Alternatively, it may be possible to use a selection of substrates in the KDG-aldolase condensation reaction to produce a wider range of compounds, as has been done with the NAL enzyme (Kim et al., 1988). As mentioned previously, the synthesis of carbon-carbon bonds is one of the most challenging aspects of synthetic chemistry. The ability of KDG-aldolase to perform this reaction in one step, whilst maintaining stereospecific integrity, is of notable interest.

KDG-aldolase from *S. solfataricus* has the additional advantage that being from a thermophilic organism, the enzyme is expected to display a high degree of thermostability. This will increase both the shelf life and the stability of the enzyme for biotransformation purposes, making it an attractive candidate for industrial use. Performing reactions at high temperatures increases the solubility of substrates, which otherwise can be a limiting factor for some biotransformation reactions. For example, when using a low specificity D-threonine aldolase to resolve L-*threo*-MTPS, an antibiotic precursor, the substrate could only be used at low concentrations in the reaction due to the solubility limitation of DL-*threo*-MTPS at the 50°C reaction temperature (Liu et al., 1999). The potential of KDG-aldolase to be active at much higher temperatures may lead to more efficient reaction processes.

Another advantage of using a thermostable enzyme expressed in a mesophilic host, is the ease with which the recombinant protein purification can be carried out, which also reduces the overall processing cost of enzyme production. Having grown cultures of *E. coli* expressing the recombinant gene, the cells can simply be broken and the desired thermostable protein significantly purified by use of a heat step followed by centrifugation to separate heat denatured *E. coli* proteins.

1.4 AIMS OF PROJECT

Clearly, there is significant potential for KDG-aldolase in the large-scale production of KDG and other carbohydrates. To realise this potential, it is necessary to obtain the enzyme in sufficient quantities for a variety of characterisation experiments. Recombinant protein, expressed in a mesophilic host, will enable the simple production of enzyme. However, as the gene for the KDG-aldolase enzyme has not yet been cloned or sequenced, there is a substantial amount of preliminary groundwork to be done. This project aims to purify the wild type enzyme from *S. solfataricus*, clone and sequence the gene, and facilitate its recombinant expression in *E. coli*. The gene sequence will be compared with other sequences in the databases to obtain information concerning similar enzymes. Characterisation of the purified protein will supplement the understanding of glucose metabolism in *S. solfataricus* and provide fundamental information concerning the reaction mechanism, thermostability and structure of the enzyme, which can be compared to other aldolase enzymes. This information will aid definition of operational parameters of the enzyme if it is to be used in large-scale biotransformations. Crystallisation of the enzyme, and a solution to the three-dimensional structure, will allow judicious predictions concerning the reaction mechanism to be made, and will enable rational modification of the substrate specificity of the enzyme. Finally, a large-scale biotransformation reaction will provide direct evidence for the potential value of this enzyme for biosynthetic purposes, and will also generate KDG for further fundamental studies.

2 MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Culture material

Tryptone for LB culture media was purchased from Amersham, Bury, England. Yeast extract and bactoagar were both from DIFCO, Detroit, Michigan, USA). Ampicillin was from Sigma-Aldrich Company Ltd., Poole, Dorset, England. X-gal was from Alexis Corporation, CH-4448 Läufelfingen, Switzerland, and IPTG from Calbiochem-Novabiochem Corporation, La Jolla, CA 92039, USA.

LB medium is made up of 1 % (w/v) bactotryptone, 0.5 % (w/v) NaCl, 0.5 % (w/v) yeast extract. LB plates are composed of LB medium with 15 g/L agar.

2.1.2 Protein purification apparatus

All FPLC columns and matrices, either pre-packed or self-assembled, were supplied by Pharmacia Biotech, Uppsala, Sweden.

Protein electrophoresis was performed using an Atto-Corp. Mini-Atto System, and Western blot transfer of protein to nitro-cellulose using a Pharmacia semi-dry transfer blotter.

2.1.3 Reagents

Tris was obtained from GibcoBRL Life Technologies Ltd., Paisley, Scotland. DMSO was from BDH Laboratory Supplies, Poole, Dorset. Protogel™ acrylamide stock solution was supplied by National Diagnostics, Atlanta, Georgia 30336, USA. Bradford reagent was either supplied by Pierce, Rockford, Illinois, USA, or prepared using the constituent reagents from the suppliers described below. Ammonium persulphate was obtained from BioRad

Laboratories Ltd., Hercules, CA, USA. Fisons Scientific equipment, Loughborough, England, or Sigma-Aldrich Company Ltd, Poole, Dorset, UK supplied all other reagents (SLR and AnalaR grades).

2.1.4 Molecular biology apparatus and reagents

PCR was carried out on a Cetus DNA Thermal Cycler (Perkin-Elmer, Norwalk CT, USA). Expand™ High Fidelity Taq polymerase and buffer were from Boehringer Mannheim, GmbH, Germany.

The MERmaid® and Geneclean® Kits for purification of oligonucleotides were obtained from Bio101, Vista, CA 92083, USA. The pGEM®-T Vector System, including T4 DNA ligase and buffer, *E. coli* JM109 cells, and the Wizard™ *Plus* Minipreps DNA Purification System were from Promega, Madison WI, USA.

SeaKem LE agarose was obtained from FMC BioProducts, Rockland, ME 04841, USA, and the 1kb ladder for agarose gels was from GibcoBRL Life Technologies Ltd., Paisley, Scotland. All restriction enzymes and their buffers are from New England BioLabs Inc., MA, USA. Deoxynucleoside 5' triphosphates (dATP, dCTP, dGTP, dTTP) were from Bioline Ltd., London, UK.

2.1.5 Computational analysis

DNA and amino acid sequence analyses were carried out using the GCG Sequence Analysis Software Package Version 7.0 (Genetics Computer Group, Inc., Madison, WI, USA).

All graphical representations of data were produced using Microsoft® Excel 98 (Microsoft Corporation, Redmond, WA, USA). Kinetic parameters were calculated using the computer programme ENZPAK version 3.0, which was supplied by BIOSOFT, Cambridge, UK.

2.2 PROTEIN PURIFICATION TECHNIQUES

2.2.1 Heat step

Cell extract was heated in 1 cm diameter glass test tubes in a water bath at 90°C for 15 min. Precipitated proteins were then removed by centrifugation at 12,000 g for 1 h at 4°C.

2.2.2 FPLC Superdex 200 gel filtration

The column was equilibrated and run in 20 mM Tris, pH 8.5. Samples were applied using a 2 ml sample loop and 1 ml fractions collected after the 20 ml void volume. The flow rate was 1 ml/min.

2.2.3 FPLC HiTrapQ anion exchange

The column was equilibrated in 20 mM Tris, pH 8.5. Samples were applied using sample loops of various sizes or using a super loop, and the column was washed in the same buffer. Protein was then eluted in 100 ml, using a gradient of 0 – 0.2 M NaCl in 20 mM Tris, pH 8.5. The flow rate was 2 ml/min. 1 ml fractions were collected.

2.3 PROTEIN ANALYTICAL TECHNIQUES

2.3.1 SDS-PAGE

SDS-polyacrylamide gel electrophoresis was carried out according to the method of Laemmli (1970). Briefly, 10% (w/v) running gel and stacking gels were composed of Protogel™ acrylamide stock containing 30% (w/v) acrylamide and 0.8% (w/v) bis-acrylamide and gel buffer concentrate and then polymerised with ammonium persulphate and TEMED according to Table 2.1 below.

Running gel buffer concentrate contained 1.5 M Tris, pH 8.9, and 0.4% (w/v) SDS. Stacking gel buffer concentrate contained 0.48 M Tris, pH 6.8, and 0.4% (w/v) SDS.

	10% Running gel	Stacking gel
Protogel (ml)	4	0.9
Gel buffer concentrate (ml)	3	2.4
H_2O (ml)	5	3.6
Ammonium persulphate 0.1% (µl)	50	50
TEMED (µl)	12.5	10

Table 2.1 SDS-PAGE components.

Samples for analysis were diluted in 2× loading buffer (0.125 M Tris, pH 6.8, 4% (w/v) SDS, 20% (w/v) sucrose, 0.08% (v/v) bromophenol blue, 10% (w/v) β-mercaptoethanol,) and heated at 100°C for 5 min.

Gels were run in 5.2 mM Tris, 0.01% (w/v) SDS and 0.04% (w/v) glycine on a Atto-Corporation Mini-Atto System according to the manufacturer's instructions at 10mA through the stacking gel and 20mA through the resolving gel, both at constant voltage.

Molecular weight standards were from BioRad (low molecular weight range): Phosphorylase b (M_r 97,400), Bovine serum albumin (M_r 66,200), Ovalbumin (M_r 45,000), Carbonic anhydrase (M_r 31,000), Soybean trypsin inhibitor (M_r 21,500) and Lysozyme (M_r 14,400).

Gels were stained in Coomassie staining solution (0.25% (w/v) Coomassie Blue R, 45% (v/v) methanol and 10% (v/v) acetic acid) for 30 - 60 min and then destained in 5% (v/v) methanol and 7.5% (v/v) acetic acid.

2.3.2 Protein concentration determination

Protein concentrations were determined by the method of Bradford (1976). Briefly, 0.1 ml protein samples of unknown concentration were incubated for 15 min at room temperature with 0.9 ml Bradford reagent (0.01% (w/v) Coomassie blue R, 5% (v/v) ethanol, 8.5% (v/v) H_3PO_4). These were then read at 595 nm against a blank of 0.1 ml H_2O in 0.9 ml

Bradford reagent. Unknown values were compared with a calibration curve of A_{595} *versus* serum albumin concentration.

2.4 DNA METHODS

2.4.1 PCR amplification of DNA

Expand™ High Fidelity Taq polymerase was used in a reaction volume of 100 µl overlaid with 20 µl mineral oil to prevent evaporation. Final concentrations in the reaction mixture were 50 µM of each deoxynucleoside 5' triphosphate (dATP, dCTP, dGTP, dTTP), primers at 0.3 mM each, approximately 100 ng DNA and 2.5 U of polymerase in a 100 µl volume of 1 × PCR reaction buffer (supplied by the manufacturer, contains $MgCl_2$). Negative controls with H_2O in place of the 5' primer, 3' primer or DNA are routinely used.

The mixture was heated to 94°C for 4 min, maintained at 85°C whilst the polymerase was added and then 30 reaction cycles were carried out using the following profile:

Denaturation	94°C	1.5 min
Annealing	48°C	1.5 min
Extension	72°C	2 min

A final 10 min extension was run after the last cycle had finished to ensure that primer extension was complete.

PCR products were visualised by electrophoreses of 15 µl of each sample though a 1% (w/v) agarose gel.

2.4.2 DNA Purification (MERmaid® and Geneclean®)

The MERmaid® and Geneclean® kits were used to purify DNA from a 1% agarose gels. The manufacturer's instructions were followed.

2.4.3 Ligation

Ligation of DNA inserts into plasmid vectors was carried out according to Promega Protocols and Applications Guide (1996). Using Equation 2.1, 10 µl reactions were set up with the appropriate volume of insert and plasmid DNA, in the molar ratios of 1:1, 3:1 and 6:1 respectively.

$$\left(\frac{vector\,(ng)\times insert\,(kb)}{vector\,(kb)}\right)\times(molar\ ratio\ insert:vector)=insert\,(ng)\quad \textbf{Equation 2.1}$$

T4 DNA ligase was used at a concentration of 1 unit per reaction (supplied at 1 unit / µl) with appropriate addition of the buffer supplied. The reaction was incubated for 3 – 16 h at 4°C.

2.4.4 Preparation of JM109 competent cells

50 ml LB media in a 500 ml flask was inoculated with several colonies of JM109 cells and grown to an OD_{600nm} of 0.4. The cells were harvested at 3,000 g for 5 min, resuspended in ice-cold 50 mM $CaCl_2$ and left on ice for 20 min. Cells were again harvested at 3,000 g for 5 min, and this time resuspended in 5 ml 50 mM $CaCl_2$ with 15 % (v/v) glycerol. 200 µl aliquots of cells were dispensed into chilled 1.5 ml tubes, and snap frozen in dry ice / methanol, before storage at -70°C until further required.

2.4.5 Transformation of *E. coli* JM109 competent cells

To 2 µl of ligation mix 50 µl of thawed *E. coli* JM109 competent cells was added and incubated on ice for 20 min. The cells were then heat-shocked at 42°C for 90 s and then cooled on ice for 2 min. A volume of 900 µl of LB medium was then added and this mixture incubated for 1 h at 37°C. Cells were then spread on LB plates supplemented with 100 µg ampicillin/ml and incubated overnight at 37°C.

2.4.6 Preparation of plasmid DNA - 'miniprep'

JM109 *E. coli* cells transformed with plasmid vector containing ligated insert were cultured overnight at 37°C in 5 ml of LB medium supplemented with 100 μg ampicillin/ml. Insert-containing plasmid DNA was prepared using the Wizard™ *Plus* Minipreps DNA Purification System according to the manufacturer's instructions.

2.4.7 DNA restriction digest

Restriction digests were set up as per the instructions from the manufacturer of the restriction enzyme. Routinely, 1 μg DNA in a 10 μl reaction was incubated for 2 h at 37°C with 1 μl (2 – 10 units) restriction enzyme, and 1 μl of the appropriate 10× buffer. If necessary, as in the case of BamH1, 10 μg of BSA was added (0.1 mg/ml). The digestion was stopped by the addition of agarose electrophoresis loading buffer to final concentrations of 7% (w/v) sucrose, 0.04% (w/v) bromophenol blue, 0.04% (w/v) xylene cyanol. Fragments were analysed by agarose gel electrophoresis.

2.5 DNA ANALYTICAL TECHNIQUES

2.5.1 Agarose gel electrophoresis

DNA was analysed by horizontal agarose gel electrophoresis. Gels were composed of 1% (w/v) agarose melted in TAE buffer (40 mM Tris acetate, 1 mM EDTA), with ethidium bromide (0.5 μg/ml) added. The gel was allowed to set in a perspex mould with comb in place to form wells. Once set, the comb was removed, the gel placed in an electrophoresis tank and covered with TAE (0.04 M Tris-acetate, 0.001 M EDTA). Samples were loaded with loading buffer (final concentrations; 7% (w/v) sucrose, 0.04% (w/v) bromophenol blue, 0.04% (w/v) xylene cyanol), and electrophoresed at a constant voltage of 50 - 90 V. DNA was visualised under UV transillumination.

2.5.2 DNA sequencing

DNA sequencing was kindly performed by Tim Marriner using a Perkin Elmer ABI PRISM™ 377 DNA sequencer (Appplied Biosystems, Fostercity CA, USA).

3 KDG-ALDOLASE ACTIVITY ASSAYS

3.1 INTRODUCTION

With any enzyme it is important to have a reliable assay for both detection and quantification of enzyme activity. Ideally, an enzyme would have a substrate, product or co-factor that absorbs light of a particular wavelength, such that the reaction can be directly followed spectrophotometrically. An example of this would be glucose dehydrogenase, an enzyme preceding KDG-aldolase in the Entner-Doudoroff pathway of *S. solfataricus* (De Rosa et al., 1984). This enzyme catalyses the conversion of glucose to gluconate, with the concomitant reduction of NAD^+ or $NADP^+$ to yield NADH or NADPH (Figure 3.1).

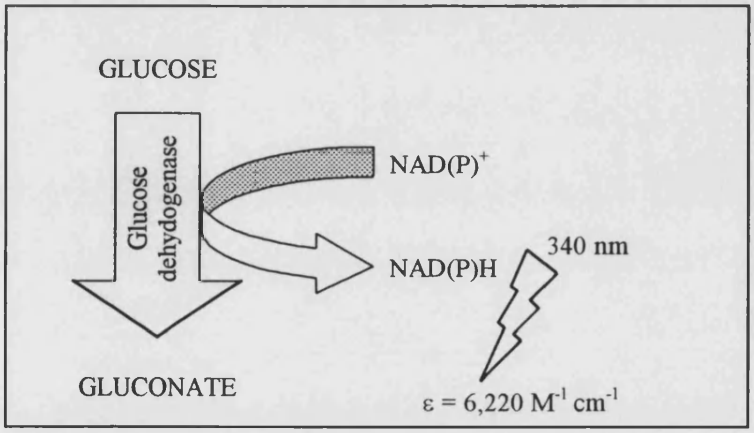


Figure 3.1 The conversion of glucose to gluconate, catalysed by glucose dehydrogenase from *S. solfataricus*. Production of NAD(P)H can be monitored by the increase in absorbance at 340 nm.

NADH absorbs light at 340 nm, and so its production can be monitored by following the increase in absorbance, and converted to the amount produced using the extinction coefficient of 6,220 M^{-1} cm^{-1}. As the conversion of glucose to gluconate and $NAD(P)^+$ to

NAD(P)H are equimolar, the number of moles of NAD(P)H produced is the same as the number of moles of gluconate produced.

It is unfortunate in the case of KDG-aldolase that neither substrates nor products absorb light in such a way, and therefore the reaction cannot be followed in so straightforward a manner. However, several attempts have been made at alternative methods for the detection and quantification of KDG and similar 2-keto-3-deoxy acids. Four such methods are detailed here, three of which were evaluated for their potential application at various points in this study.

3.1.1 Thiobarbituric acid assay

This complex method of detecting 2-keto-3-deoxy acids was initially devised by Weissbach and Hurwitz (1959) and has subsequently been amended for varying uses by several groups, including Bender and Gottschalk (1973) who used it for the determination of KDG. The assay procedure, illustrated schematically in Figure 3.2, involves the oxidation of KDG with sodium periodate to produce formyl pyruvic acid. This oxidation reaction is then terminated by the addition of sodium arsenite, and the product treated with thiobarbituric acid to produce a pink chromophore that can be stabilised and intensified by the addition of DMSO (Skoza and Mohos, 1976). Its absorbance can be measured at the wavelength 549 nm and the amount of KDG calculated using a molar extinction coefficient of $67.8 \times 10^3\ M^{-1}\ cm^{-1}$ (Bender and Gottschalk, 1973).

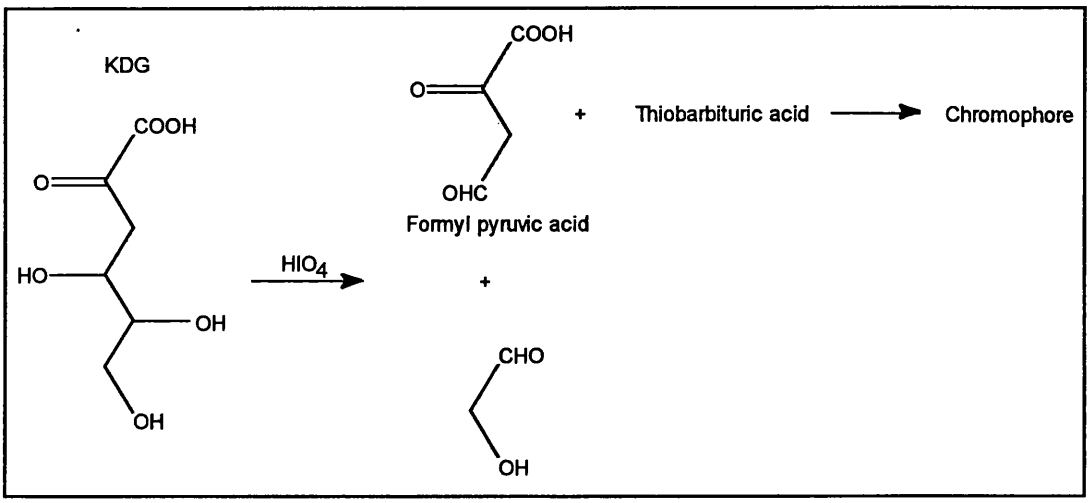


Figure 3.2 Schematic representation of the thiobarbituric acid assay.

Due to the nature of the detection method it is clear that the formyl pyruvate derivative is required for colour development, and therefore pyruvate analogues that do not form such a derivative may be unsuitable for testing in this assay. It is also evident from Figure 3.2 that formyl pyruvic acid is only generated if there are hydroxyl groups on C_4 and C_5 of the KDG-aldolase reaction product. As the C_4 hydroxyl group is generated from the aldehyde, only α-hydroxyaldehydes will give a C_4, C_5 diol, and therefore this assay will also limit the number of alternative aldehyde substrates that can be tested.

3.1.2 Pyruvate depletion assay

If the reaction catalysed by KDG-aldolase is stopped after a certain time, the amount of unconverted pyruvate, or the amount of pyruvate produced, can be determined by using a second reaction with the enzyme lactate dehydrogenase. This enzyme will convert pyruvate to lactate, with the concomitant oxidation of NADH to NAD^+ (Figure 3.3).

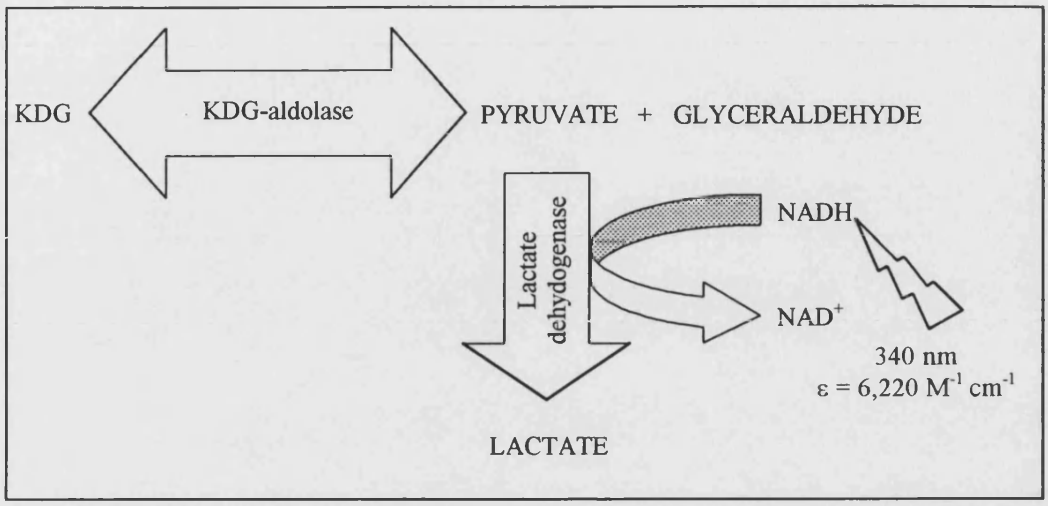


Figure 3.3 Schematic diagram illustrating the pyruvate depletion assay. The KDG-aldolase reaction followed by lactate dehydrogenase conversion of pyruvate to lactate, with concomitant oxidation of NADH. Disappearance of NADH can be monitored by the decrease in absorbance at 340 nm.

The pyruvate measured in the lactate dehydrogenase assay can be either the pyruvate remaining after the aldol condensation reaction with production of KDG, or the amount of pyruvate produced by the retro-aldol reaction. A disadvantage of this assay method when measuring KDG-aldolase activity in the condensation direction is that the decrease in pyruvate would be relatively small compared with the initial concentration. This potentially introduces a large error into the calculation of enzyme activity. In this direction a discontinuous method must be used. If used in the retro-aldol direction, starting with KDG, the pyruvate produced can be measured continuously using a thermostable lactate dehydrogenase in a coupled reaction, as long as neither the LDH enzyme nor the NADH becomes limiting.

This assay was used for the substrate specificity experiments in section 7.3.7 as some of the alternative aldehyde substrates were found to interfere with the TBA assay. It would not be possible to use this assay for testing substrates alternative to pyruvate.

3.1.3 RI-HPLC detection method

Prior calibration of the RI-HPLC output for substrate/product concentrations allows the simple quantification of substrates and products at any stage during the KDG-aldolase reaction. Theoretically this method should be ideal for the analysis of potential alternative substrates for the KDG-aldolase enzyme, provided that an appropriate column is used for elution of the substrates/products. A disadvantage of this method is the run time (25 min) of each sample to be analysed; therefore, it would not be ideal for the routine analysis of enzyme activity, for instance when locating the enzyme-containing fractions following a chromatography purification step.

3.1.4 The semicarbazide method

The semicarbazide method of detecting KDG was developed by MacGee and Doudoroff (1954) and later modified by Kersters et al. (1971). The assay method involves the reaction of semicarbazide with the α-keto acid to give the semicarbazone derivative of the compound. Unfortunately, both KDG and pyruvate are α-keto acids and so this assay method is not viable for analysing KDG-aldolase activity, and will not be discussed further.

3.2 MATERIALS AND METHODS

3.2.1 Materials

KDG was obtained from Eric J. Toone, Department of Chemistry, Duke University, Durham, North Carolina, USA. Absorbances were measured using a Perkin Elmer Lambda Bio uv/vis spectrophotometer (Perkin-Elmer, Beaconsfield, Bucks., UK). HPLC apparatus and refractive index detector were from BioRad, York, North Yorkshire, UK.

Rabbit muscle lactate dehydrogenase was from Boehringer Mannheim, Lewes, East Sussex, UK. All other chemicals were supplied by Sigma-Aldrich Company Ltd., Poole, Dorset, UK.

3.2.2 Thiobarbituric acid enzyme assay

Aldolase activity was measured using a modification of the thiobarbituric acid assay developed by Skoza and Mohos (1976) and Gottschalk and Bender (1982). Reactions of total volume 250 µl were incubated at 70°C in 50 mM sodium phosphate buffer, pH 6.0, with 50 mM pyruvate and 20 mM DL-glyceraldehyde and an appropriate volume of cell extract or purified enzyme between 10 and 50 µl. After 10 min, 100 µl samples were removed and the reaction stopped by addition of 10 µl of 12% TCA acid. Precipitated proteins were then removed by centrifugation. 50 µl of this sample containing the aldolase product was then oxidised by addition of 125 µl of 25 mM periodic acid / 0.125 N H_2SO_4 at room temperature for 20 min. Oxidation was terminated by addition of 250 µl of 2% (w/v) sodium arsenite / 0.5 N HCl. 1 ml of 0.3% (w/v) thiobarbituric acid was then added and the chromophore developed by heating at 100°C for 10 min. A sample of this was then removed and the colour intensified by adding to an equal volume of DMSO. Absorbance was read at 549 nm. The molar absorbance coefficient for the chromophore used was 67.8×10^3 M^{-1} cm^{-1} (Skoza and Mohos, 1976; Gottschalk and Bender, 1982).

3.2.3 Pyruvate depletion enzyme assay

An initial KDG-aldolase catalysed reaction was carried out in a similar fashion to that for the thiobarbituric acid assay (section 3.2.2). At appropriate times, 100 µl samples were removed and the reaction stopped by addition of 10 µl of 12% TCA acid. Precipitated proteins were then removed by centrifugation. A sample of the stopped reaction was diluted appropriately in 100 mM sodium phosphate buffer, pH 7.5, such that the pyruvate content in 20 µl would be less than 0.2 µmoles. 20 µl of this diluted sample was added to 980 µl of 100 mM sodium phosphate buffer, pH 7.5, containing 0.2 mM NADH and 25 µg/ml rabbit muscle lactate dehydrogenase, and incubated for 30 min at room temperature to allow complete conversion of pyruvate to lactate, and NADH to NAD^+. The absorbance was measured at 340 nm, against a blank where the diluted sample of stopped KDG-aldolase reaction was replaced with water. The decrease in absorbance is converted to pyruvate remaining (M), using the molar absorbance coefficient for NADH of 6,220 M^{-1} cm^{-1}.

3.2.4 Chemical analysis by refractive-index high performance liquid chromatography (RI-HPLC)

0.5 ml samples from a KDG-aldolase catalysed reaction were diluted 1:1 with 0.5 ml 8 mM H_2SO_4 and 5 µl loaded onto an ion exchange column (BIO-RAD HPLC Organic Acid Analysis Column Aminex Ion Exclusion HPX-87H, 300 × 7.8 mm [column 76, Serial no. 77999]), linked to a Hewlett Packard Series 1100 HPLC with refractive index detector (Hewlett Packard 1047A). The mobile phase used was 8 mM H_2SO_4. Output is by graphical representation of compounds as they are eluted, with time of elution on the x-axis, and refractive index of the compound on the y-axis.

3.2.5 Calculation of standard error from standard deviation

Using the standard deviation calculated by the computer package Microsoft Excel, the standard error can be calculated according to Equation 3.1.

$$\text{Standard error} = \frac{\text{Standard deviation}}{\sqrt{\text{number of replicates}}} \qquad \textbf{Equation 3.1}$$

3.2.6 Combining errors

Two or more terms with standard errors can be combined using Equations 3.2 and 3.3 to give the overall value with the revised standard error.

$$Z = A + B \qquad \textbf{Equation 3.2}$$

$$(\Delta Z)^2 = (\Delta A)^2 + (\Delta B)^2 \qquad \textbf{Equation 3.3}$$

where A and B are the terms to be combined, Z is the sum of A and B, and ΔA, ΔB and ΔZ are their associated standard errors.

Procedures used in this chapter that are not mentioned in this section will be found in chapter 2.

3.3 RESULTS

3.3.1 Thiobarbituric acid assay

The ease of manipulation of the thiobarbituric acid assay, and the standard errors involved, were evaluated by analysis of the rate of production of KDG over time by the KDG-aldolase enzyme with the substrates Na-pyruvate and DL-glyceraldehyde. Graphical representations of results from two experiments are shown in Figure 3.4.

Individual enzyme activities from two replicate experiments were 0.76 (± 0.017) μmoles/min/ml (Figure 3.4a) and 0.77 (± 0.014) μmoles/min/ml. The overall percentage standard error was calculated to be ± 1.5 %.

3.3.2 Pyruvate depletion assay

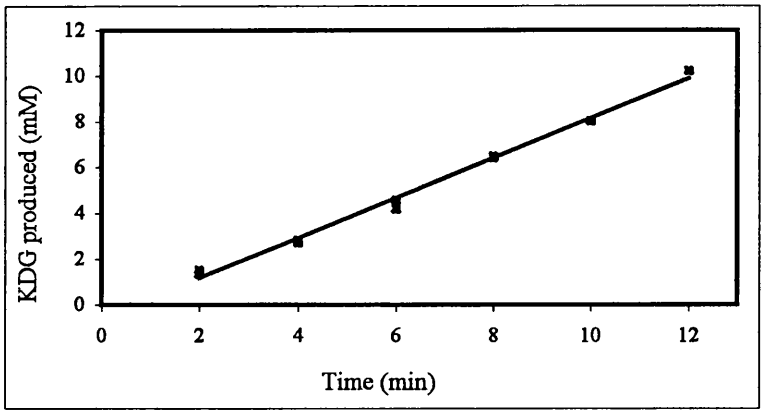
The pyruvate depletion assay was analysed in a similar manner to the thiobarbituric acid assay (Figure 3.5). A comparison of the two methods was made (Figure 3.6) by testing samples from a single enzyme reaction in both assays. The pyruvate depletion assay gave an activity of 16.77 (± 0.94) μmoles/min/ml, whilst in the TBA assay the same samples gave an activity of 22.58 (± 0.45) μmoles/min/ml. The percentage standard error of the pyruvate depletion assay was ± 5.6 %. With the same samples tested in the TBA assay the error was ± 2.0 %.

3.3.3 RI-HPLC detection method[1]

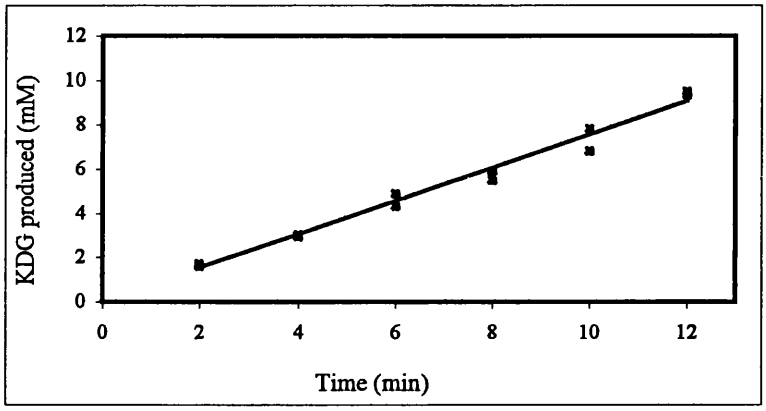
3.3.3.1 KDG quantification

Due to the hygroscopic nature of KDG, it was necessary to determine the purity of the KDG standard to be used. This was established using the detection procedure outlined for the

[1] Some details given in this section are repeated for the reader's convenience in chapter 9, KDG Biotransformation.



(a)



(b)

Figure 3.4 Analysis of standard error in the thiobarbituric acid assay. Graphical representations of two assays are illustrated, with enzyme activities 0.76 (± 0.017) μmoles/min/ml (a) and 0.77 (± 0.014) μmoles/min/ml (b).

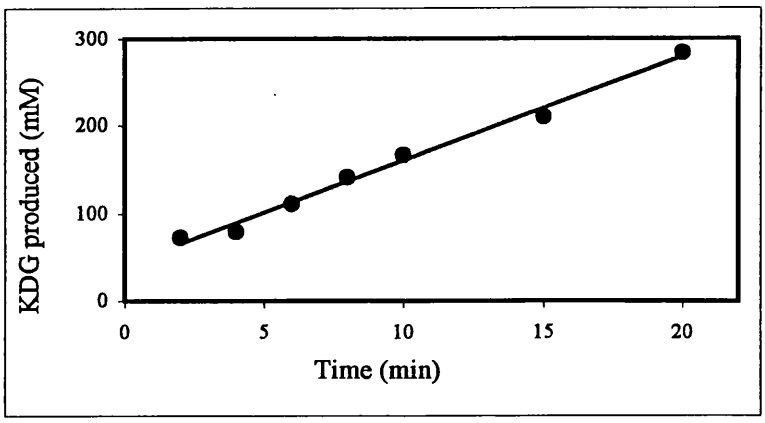


Figure 3.5 Analysis of standard error in the pyruvate depletion assay. KDG-aldolase activity is 16.77 (± 0.94) μmoles/min/ml.

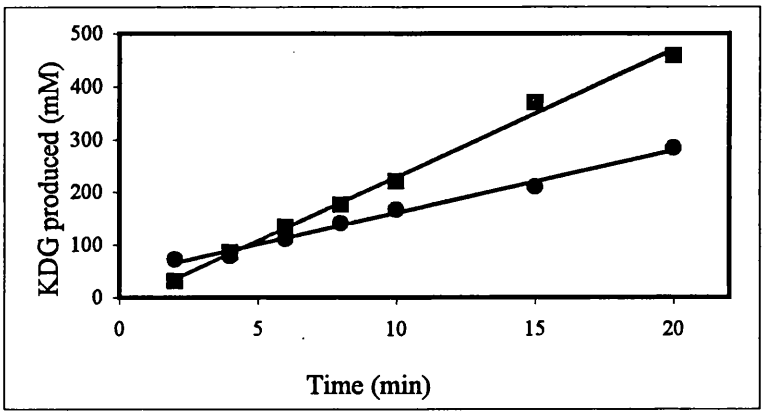


Figure 3.6 Comparison of the thiobarbituric acid assay (■) and the pyruvate depletion assay (●). KDG-aldolase activity measured in the thiobarbituric acid assay is 22.58 (± 0.45) μmoles/min/ml and in the pyruvate depletion assay (also shown in Figure 3.5) is 16.77 (± 0.94) μmoles/min/ml.

thiobarbituric acid assay. It was ascertained that the standard contained 38 % (w/w) KDG. The additional weight is believed to be due to associated water, since analysis of the KDG by RI-HPLC resulted in only two peaks with elution times identical to those for the two peaks formed by the aldol condensation reaction (section 3.3.3.2). Additional peaks would be expected if there were other contaminating compounds. All further references to amount or concentration of the KDG standard are corrected accordingly.

3.3.3.2 RI-HPLC calibration

Elution of compounds from ion exchange.HPLC occurs at times specific to the chemical nature of the compound loaded. Detection by changes in refractive index permit retention times to be assigned to single compounds to enable identification of compounds in mixtures. Table 3.1 gives retention times for DL-glyceraldehyde, Na-pyruvate, and KDG.

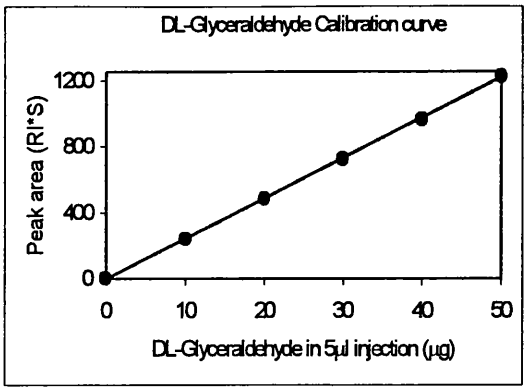
Compound	Retention time (min)
DL-Glyceraldehyde	13.7
Na-pyruvate	12.2
KDG	10.5 & 11.1

Table 3.1 HPLC retention times for DL-glyceraldehyde, Na-pyruvate, and KDG.

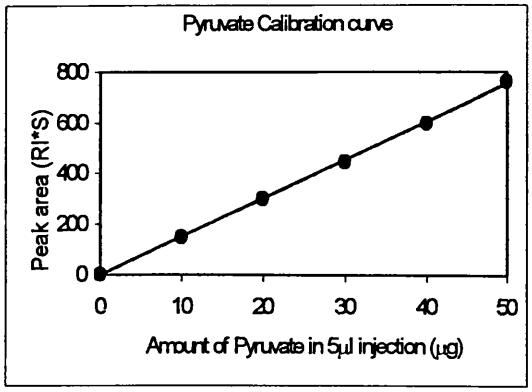
The RI-HPLC output gives a peak area proportional to the amount of compound injected. Having injected known amounts of DL-glyceraldehyde, Na-pyruvate and KDG, calibration curves of peak area versus amount of compound were plotted (Figure 3.7). The Equations of the lines (Equations 3.4 – 3.6) can be used to calculate unknown amounts of these compounds.

3.3.3.3 Enzyme assay

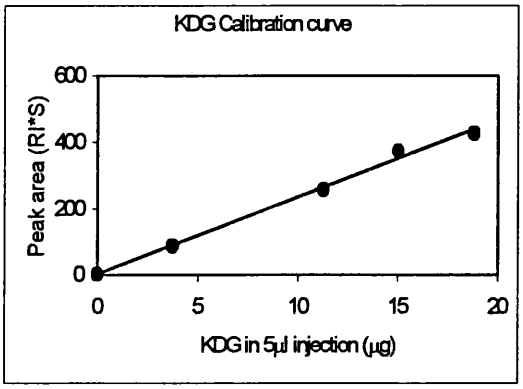
KDG-aldolase enzyme reactions were followed by periodically removing 0.5ml samples, stopping by 1:1 dilution with 8mM H_2SO_4, and analysis by RI-HPLC. Peaks for each substrate / product are seen, with the area of the peak proportional to the amount of



(a)



(b)



(c)

Figure 3.7 Calibration curves of RI-HPLC peak area against amount (µg) of DL-glyceraldehyde (a), Na-pyruvate (b) and KDG (c) in 5µl injection peak.

Equation 3.4 – 3.6

amount (µg) of DL-glyceraldehyde in 5µl sample = (peak area + 3.2548)/24.329 **3.4**

amount (µg) of Na-pyruvate in 5µl sample = (peak area + 2.1552)/15.134 **3.5**

amount (µg) of KDG in 5µl sample = (peak area + 0.5053)/23.339 **3.6**

substrate / product in the sample injected. Using the Equations 3.4 – 3.6, the progress of the reaction was followed by calculating the amounts of substrate / product at each time point. Scheme 3.1 gives an example calculation. Table 3.2 shows the results for each substrate / product and Figure 3.8 illustrates these figures in graphical form. The production of KDG over time mirrors the depletion of both Na-pyruvate and DL-glyceraldehyde.

The rate of production of KDG over the linear range of the reaction is 1.35 (± 0.05) μmoles/min/ml and the percentage standard error is ± 3.7 %. The rates of consumption of DL-glyceraldehyde and pyruvate over the linear range of the reaction are 1.42 (± 0.05) and 1.26 (± 0.04) μmoles/min/ml respectively and the percentage standard errors are ± 3.7 % and ± 3.4 % respectively.

Scheme 3.1 Example calculation for converting RI-HPLC peak area to the concentration of Na-pyruvate in reaction:

Na-pyruvate in 5μl injection (μg)	=	(peak area + 2.1552) / 15.134
Na-pyruvate in sample injected (μg/μl)	=	(μg Na-pyruvate in 5μl injection)/5
Na-pyruvate in sample injected (mg/ml)	=	Na-pyruvate in sample injected (μg/μl)

Sample was diluted 1:1 in 8mM H_2SO_4;

Na-pyruvate in reaction (mg/ml)	=	mg/ml Na-pyruvate in sample injected x 2

M_r of Na-pyruvate is 110;

[Na-pyruvate] (M)	=	mg/ml Na-pyruvate in reaction ÷ 110

Time	Concentration in reaction (mM)		
(min)	DL-Glyceraldehyde (O)	Na-Pyruvate (Δ)	KDG (■)
0	44	48	
2	41	45	
6	36	41	9
8	33	38	11
10	31	36	14
15	25	31	19
20	20	26	24
25	15	22	31
30	12	18	32
60		12	34

Table 3.2 KDG-aldolase reaction progress. As the concentrations of substrates decrease there is a concomitant increase in KDG.

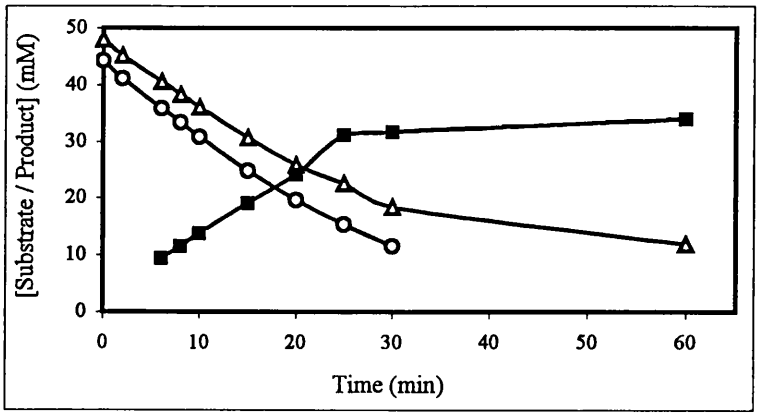


Figure 3.8 Graph illustrating KDG-aldolase reaction progress. The production of KDG (■) and consumption of DL-glyceraldehyde (O) and Na-pyruvate (Δ) are shown.

3.4 DISCUSSION

Of the four potential enzyme assays for KDG-aldolase, three have been analysed in this chapter for use throughout this study.

3.4.1 Requirements for enzyme assay

To enable the purification of the KDG-aldolase enzyme from *S. solfataricus*, or recombinant enzyme from *E. coli* total cellular protein, it is necessary to have an assay procedure to identify the presence of the enzyme, and also to quantify the amount of activity. In this way, fractions following each purification step that do not contain KDG-aldolase activity can be eliminated and those that do contain activity can be retained. As there are sometimes large numbers of fractions to be analysed before the next purification step can begin, for instance following column chromatography, an assay that is not overly time consuming is of importance.

The characterisation of the enzyme for its thermostability and optimum temperature of activity similarly requires an assay that will accurately quantify the enzyme activity. Kinetic analyses (i.e. determination of K_m and V_{max}) likewise require a sensitive assay that can detect small changes in substrate / product concentrations.

Analysis of the substrate specificity of the enzyme is complex, as the assay must be able to detect either the product formation or substrate depletion, or both, for a wide variety of compounds. For assay methods where detection is based on chemical modification this would not be possible.

Finally, the ability to follow the course of a large-scale biotransformation reaction accurately is essential to any scale up process. For this purpose it would be preferable to make use of a discontinuous assay technique that allows identification of both the product and substrate concentrations at various time points, without severely depleting the reaction volume.

3.4.2 Ease of use and repeatability of assay

The thiobarbituric acid assay has the disadvantage of being discontinuous, such that the reaction has to be stopped and the product formed then quantified. This makes the analysis of one sample time consuming, but as a large number of samples can be analysed simultaneously this would not generally be a handicap. A significant advantage of the assay is its reproducibility, with a standard error of 1.5 %. This assay is most profitable for assaying in the direction of aldol condensation, measuring KDG produced, as the substrates pyruvate and glyceraldehyde are readily available in pure form and at low cost.

The pyruvate depletion assay is also a discontinuous assay when measuring the rate of the aldol condensation reaction. Following reaction of pyruvate and glyceraldehyde to form KDG, the reaction is stopped and the pyruvate remaining converted to lactate by the enzyme lactate dehydrogenase. This requires incubation for a period of time sufficient for complete conversion to lactate, otherwise a falsely low result would be returned. This time disadvantage is compounded by the overall inaccuracy of the assay (standard error 5.6 %), as demonstrated in section 3.3.2. This inaccuracy is given by the necessity to measure small decreases in the pyruvate concentration.

Using the basis of the lactate dehydrogenase assay in a continuous coupled reaction to measure pyruvate formation in a retro aldol reaction has potential advantages over either the thiobarbituric acid assay or the pyruvate depletion assay as described above. This method would significantly reduce the time taken to assay each sample. The error would also be diminished, both due to a reduction in number of reactant manipulations and as the change in pyruvate concentration would be relative to zero. However, the assay would require a thermostable lactate dehydrogenase, such as from *Bacillus stearothermophilus*. Furthermore, a prominent disadvantage of this assay is the non-availability of KDG on the commercial market, and the inaccuracy involved in handling this compound due to its hygroscopic nature.

Analysis of the KDG-aldolase reaction by RI-HPLC has the distinct advantage over the other methods in that it gives an indication of all substrate / product concentrations in the reaction mixture, and when used on a larger scale than studied here would facilitate their isolation for further characterisation. The determination of compound concentrations has a high level of accuracy, with a standard error of 2.4 %. For the complete elution of Na-pyruvate, DL-glyceraldehyde and KDG a run time of 25 min is required, making this a laborious assay for routine enzyme detection and analysis. When testing alternative substrates this method also has drawbacks; not metabolites all elute from the column under the standard conditions used, and the product elution time and peak area relationship to concentration would not be known.

3.4.3 Applicability of assays

Having looked at the requirements of this study for enzyme assays, and the operational parameters of the assays available, it is possible to make informed decisions as to which assay is most applicable to which requirement.

The thiobarbituric acid assay, with its high level of accuracy and ability to screen a large number of samples simultaneously, is suitable for the majority of routine enzyme assays. This would include enzyme detection and quantification, and determination of kinetic parameters, thermostability and temperature optimum of activity. This assay would not be suitable for extensive substrate specificity analysis due to the limitations of the detection method.

The pyruvate depletion assay, however, extends the range of substrates that can be analysed to analogues of glyceraldehyde. As pyruvate is an absolute requirement of this assay, analogues of pyruvate cannot be tested. Due to the limited accuracy of the assay, this method would not be useful for other experiments.

The RI-HPLC assay method appears to be ideal for substrate specificity analysis, although investigation of suitable separation columns and elution buffers may be required for

a large range of substrate analogues. The time constraint of the assay means that it is not useful for routine enzyme activity analysis.

4 PURIFICATION OF KDG-ALDOLASE FROM *S. SOLFATARICUS*

4.1 INTRODUCTION

The primary aim of the work reported in this chapter was to purify KDG-aldolase from *Sulfolobus solfataricus* to enable an N-terminal amino acid sequence to be determined, and hence the gene to be cloned. Following this the enzyme can be recombinantly over-expressed for characterisation studies.

There is currently no reported purification schedule for a KDG-aldolase. Due to the diversity amongst aldolases it is expected that KDG-aldolase purification will not follow any previously documented protocol and therefore a novel approach is required.

Various chromatographic steps were attempted with *S. solfataricus* cell extracts for the purification of KDG-aldolase. Identification of fractions containing the enzyme of interest was carried out using the TBA assay (a detailed description of which appears in section 6). Fractions were also examined by SDS-PAGE, although the KDG-aldolase band could not be identified, as the size was unknown at this stage.

4.2 MATERIALS AND METHODS

4.2.1 Strains

Sulfolobus solfataricus P1 (DSM 1616) and MT4 cell pastes were kindly donated by Prof. R. Sharp (Centre for Applied Microbiological Research, Porton Down, UK).

4.2.2 Materials and reagents

All FPLC columns and matrices, either pre-packed or self-assembled, were supplied by Pharmacia Biotech, Uppsala, Sweden. Hydroxylapatite Bio-Gel HT was obtained from BioRad Laboratories Ltd., Hercules, CA, USA. The Piksi-H screening kit was from Affinity Chromatography Ltd, Freeport, Ballasalla, Isle of Man, British Isles.

4.2.3 Preparation of cell extracts from *Sulfolobus solfataricus*

Cell pastes of various wet weights were re-suspended to approximately 0.2 g/ml in 20mM Tris-HCl, pH 8.5, 1 mM PMSF, 1 mM EDTA. Cells were broken by five 30s bursts of sonication on ice with a 3 mm probe, 9 – 12 μm peak to peak, using a MSE 150 watt Ultrasonic Disintegrator Mk 2 (MSE, Crawley, UK).

Cell extract was obtained by centrifugation at 12,000 g for 1 h at 4°C. The pellet was discarded.

4.2.4 Ammonium sulphate precipitation

The appropriate amount of pulverised ammonium sulphate, according to Table 4.1 below (adapted from Harris and Angal [1989]), was added to the cell extract whilst stirring at 4°C and the suspension was maintained at this temperature for 15 min. Precipitated proteins were then removed by centrifugation at 12,000 g, for 20 min at 4°C. The concentration of ammonium sulphate in the supernatant was then raised to the next concentration by addition of more ammonium sulphate (see Table 4.1). Precipitated proteins were then once again separated from those not precipitated by centrifugation under the same conditions as previously, and the pellet re-suspended in twice the volume of 20 mM Tris-HCl, pH 8.5.

	Final concentration of ammonium sulphate, % saturation at 0°C		
	45%	65%	85%
Initial concentration of ammonium sulphate	g solid ammonium sulphate to add to 100ml of solution		
0%	26.2	40.4	56.7
45%	0	12.5	26.7
65%		0	13.4

Table 4.1 Ammonium sulphate precipitation. The Table indicates the amount of ammonium sulphate required to add to 100 ml solution to raise the concentration to the required level (adapted from Harris and Angel [1989]).

4.2.5 FPLC Sephacryl 300 gel filtration

An XK 50 column (5 cm diameter) of 44 cm height was packed under continuous pressure with Sephacryl 300 matrix and equilibrated with 20 mM Tris-HCl, pH 8.5. Samples were loaded using a superloop, and 5 ml fractions collected after the 200 ml void volume.

4.2.6 Hydroxylapatite chromatography

A 2.6 cm × 2 cm column of hydroxylapatite Bio-Gel HT was equilibrated with potassium phosphate buffer, pH 7.0, of various concentrations from 10 - 50 mM. The sample was applied and the column washed of unbound protein in equilibration buffer. Bound protein was eluted at 0.4 ml/min with a 500 ml phosphate gradient (10 / 50 mM - 500 mM, pH 7.0) and fractions collected.

4.2.7 Hydrophobic interaction chromatography

The Picksi-H screening kit was used according to the manufacturer's instructions. All 10 columns were equilibrated with 50 mM sodium phosphate buffer, pH 7.0, containing 1.0 M ammonium sulphate. Protein samples were applied and the columns washed of unbound

protein. Bound proteins were then eluted with 50 mM sodium phosphate buffer, pH 7.0, and fractions of unbound and bound protein tested for aldolase activity.

4.2.8 Native PAGE

Native polyacrylamide electrophoresis was performed as SDS-PAGE (section 2.3.1) but omitting SDS from loading buffer and gel buffer concentrates.

4.2.9 Preparative electrophoresis

Approximately 15 μg protein was electrophoresed though each of three parallel lanes in a native polyacrylamide gel. One lane was removed and stained with Coomassie. Without staining, 1 mm horizontal sections were cut from the remaining two thirds of the gel. These sections were then divided into two fractions. One fraction of each was crushed in 50 mM sodium phosphate buffer, pH 6.0, and assayed for KDG-aldolase activity in the TBA assay. The second fractions were crushed in 2 × loading buffer and electrophoresed by the SDS-PAGE method.

4.2.10 Protein transfer to nitrocellulose

SDS-PAGE gels, nitro-cellulose filters and filter paper were thoroughly soaked in transfer buffer (30 mM glycine, 48 mM Tris-HCl, 0.0375% (w/v) SDS, 20% methanol). A sandwich was then formed of filter paper, nitro-cellulose, gel, and filter paper and placed in a semi-dry transfer blotter. The transfer was performed at 0.8 mA/cm^2 for 1 h 15 min. Visualisation of the protein was by staining with 0.01% (w/v) Coomassie Blue R and 50% (v/v) methanol for 10 s, then destaining in 50% (v/v) methanol until the bands became apparent.

4.2.11 Protein microsequencing

Protein sequencing was kindly performed by Mrs. Janice Young at Zeneca Pharmaceuticals using an Applied Biosystems 477 Sequencer (modified to run blot cycles) with a gas phase TFA delivery.

Procedures used in this chapter that are not mentioned in this section will be found in chapter 2.

4.3 RESULTS

A summary of the purification of KDG-aldolase from *S. solfataricus* is given in Table 4.2 with corresponding SDS-PAGE analysis in fig. 4.1.

4.3.1 Preparation of soluble cell extract

Cell lysis was achieved by sonication followed by centrifugation to remove cell debris. Repeated sonication of the resuspended pellet yielded little further KDG-aldolase activity, showing that the initial sonication was sufficient for total cell lysis, and that the enzyme is soluble.

4.3.2 Heat precipitation step

Although a heat step is more commonly used to purify heat stable recombinant proteins expressed in a mesophilic host organism, it can be useful when the target protein has a higher thermostability compared to the other proteins in the cell. In this instance a heat step of 15 min at 90°C gives an 87 % yield and 3 % increase in purity, relative to the soluble cell extract fraction.

4.3.3 Ammonium sulphate precipitation

Addition of 65 % ammonium sulphate facilitates the 10-fold purification of KDG aldolase by decreasing the solubility of proteins in solution such that they precipitate or 'salt-out'. Remaining soluble proteins, including KDG-aldolase, were recovered by centrifugation.

4.3.4 Sephacryl S-300 HR gel filtration

The sample containing ammonium sulphate was loaded onto a Sephacryl S-300 HR gel filtration column to both remove the salt from the sample and to separate proteins of different M_r. This step yielded only 25 % of the activity loaded onto the column but the purity increased 2.3 times.

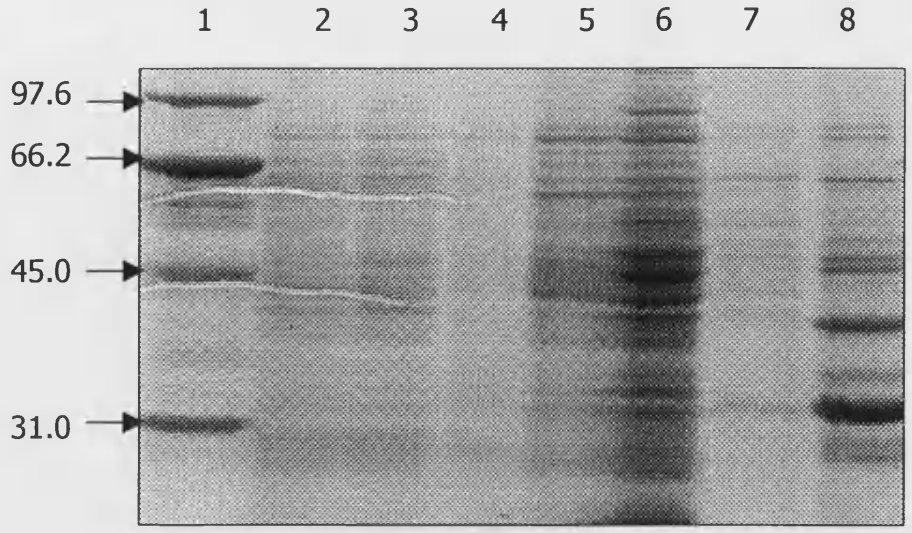


Figure 4.1 SDS-PAGE of the purification of the KDG-aldolase. Samples are as follows: lane 1, SDS-PAGE molecular weight standards; lane 2, total cell homogenate; lane 3, soluble cell extract; lane 4, supernatant after heat step and centrifugation; lane 5, 40% ammonium sulphate cut supernatant; lane 6, 65% ammonium sulphate cut supernatant; lane 7, Sephacryl S-300 HR gel filtration eluent; lane 8, HiTrap™ Q anion exchange 'best fraction'. Molecular weights indicated by arrows are in kDa.

Purification step	Total protein (mg)	Total activity (units)	Specific activity (units/mg)	Yield (%)	Purification factor (× fold)
Total cell homogenate (2)	1418	34.7	0.0245	100	1
Soluble cell extract (3)	1224	32.5	0.0265	93	1.08
90°C Heat step (4)	1036	28.1	0.0272	81	1.11
65% ammonium sulphate cut supernatant (6)	70.6	16.6	0.235	48	9.6
Sephacryl S-300 HR gel filtration (7)	7.4	4.1	0.552	12	23
HiTrap™ Q anion exchange 'best' fraction (8)	0.3	0.50	1.94	1.4	79

Table 4.2 Purification of KDG-aldolase from *S. solfataricus*. Samples after each step were analysed for activity in the TBA assay and by SDS-PAGE. Numbers in brackets after the purification step indicate the SDS-PAGE lane number in Figure 4.1. Yields and fold purification are relative to the total activity present in the total cell homogenate.

4.3.5 HiTrap™ Q anion exchange

Fractions containing the KDG-aldolase activity from the gel filtration column were pooled and loaded onto anion exchange resin. To facilitate the complete purification of the KDG-aldolase only the fraction with the highest specific activity was retained (HiTrap™ Q anion exchange 'best' fraction [Figure 4.2a]). The shoulder fractions were discarded, hence the tabulated yield of KDG-aldolase at this point is low.

4.3.6 Preparative native polyacrylamide gel electrophoresis

Approximately 15 μg of HiTrap™ Q anion exchange 'best' fraction was electrophoresed through 3 lanes in a native polyacrylamide gel. One lane was stained (Figure 4.2). The unstained lanes were sectioned and treated as described (4.2.9). The results of the TBA assay for KDG-aldolase activity indicate that sections corresponding to fractions loaded in lanes 8 and 9 of the SDS-PAGE contained activity at 0.03 and 0.04 μmoles/min/ml respectively (Figure 4.3).

4.3.7 Determination of subunit M_r by SDS PAGE

Having identified the fractions from preparative electrophoresis that contain KDG-aldolase activity, the enzyme could be identified by SDS-PAGE. KDG-aldolase runs as a single band on SDS-PAGE and is therefore either a monomer, or a homooligomeric protein. The subunit M_r of KDG-aldolase is estimated to be 32,000, by comparison to molecular weight standards.

4.3.8 N-terminal sequencing of *S. solfataricus* KDG-aldolase

The 32 kDa band corresponding to the fraction containing aldolase activity was blotted onto PVDF membrane, excised and sequenced. The 38 amino acid N-terminal sequence determined is shown below.

NH_2 PEIIT PIITP FTKDN RIDKE KLKIH AENLI RKGID KLF... CO_2H

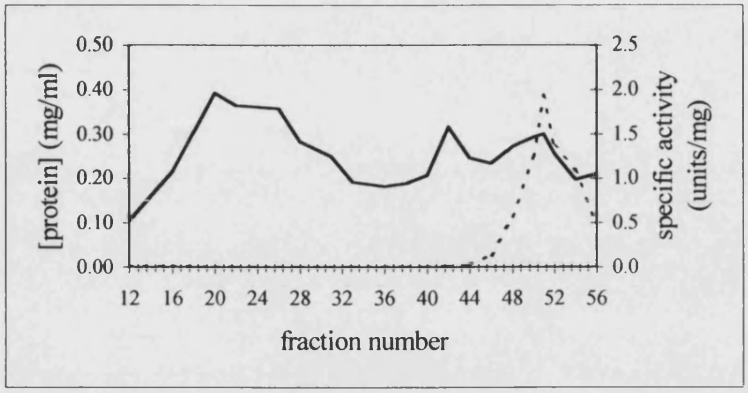


Figure 4.2a HiTrap™ Q anion exchange chromatogram. The solid line represents protein concentration and the dotted line represents KDG-aldolase specific activity. The NaCl concentration between fractions 12 and 56 varied from 40 to 130 mM linearly. Fraction 51 had the highest specific activity (HiTrap™ Q anion exchange 'best' fraction) and was retained.

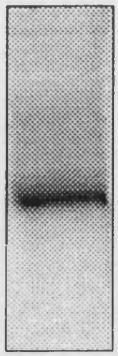


Figure 4.2 Native PAGE of anion exchange 'best' fraction. The figure shows the stained portion of the native gel. The 1 mm fractions cut correspond approximately to the region containing protein and slightly beyond.

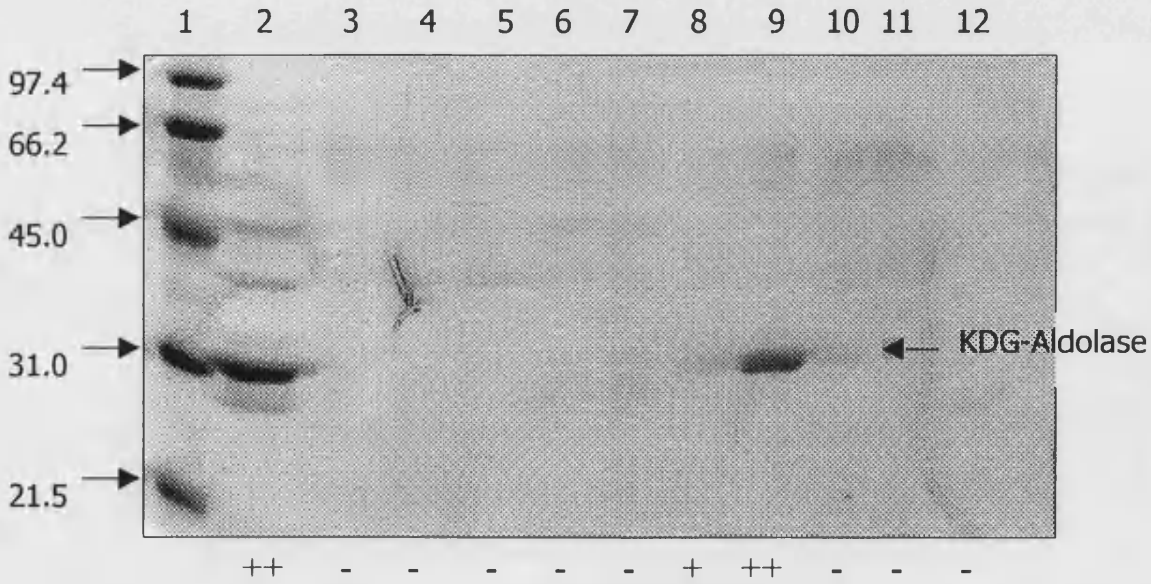


Figure 4.3 SDS-PAGE of native PAGE fractions. Samples are as follows; lane 1, SDS-PAGE molecular weight standards; lane 2, HiTrapQ anion exchange 'best fraction' as loaded on native-PAGE; lanes 3 - 12, consecutive 1 mm sections from native-PAGE. Relative KDG-aldolase activities for each fraction are illustrated by -/+/++ under the gel.

The N-terminal sequence was compared with all other sequences in the SwissProt database and gave 34 % and 30 % identities respectively to the N-terminal regions of fructose bisphosphate aldolase from *Staphylococcus carnosus* and *N*-acetylneuraminate lyase subunit from *E. coli*.

Other purification techniques were attempted with limited success. Hydroxylapatite chromatography proved to be of no use as aldolase activity was not recoverable on several occasions, although it was evident that protein had bound to the column and had subsequently been eluted. The Picksi-H kit was used to screen potential hydrophobic interaction columns as either positive or negative purification steps. Quantification of aldolase activity in unbound and eluted fractions for each column were unreliable and gave inconsistent results on every occasion.

4.4 DISCUSSION

To enable extensive characterisation studies of the KDG-aldolase enzyme, and to provide sufficient quantities with which to perform large-scale biotransformations, it is necessary to obtain KDG-aldolase in large quantities by recombinant expression. To do so, the KDG-aldolase gene is required. This chapter has concerned itself with the purification of the wild type enzyme from the host organism to enable sequencing of the N-terminus. Subsequently, this amino acid sequence will allow the design of DNA probes with which to probe a λ library to obtain the gene sequence.

It had been anticipated that KDG-aldolase would not have great similarity to other aldolases previously purified due to its distinction of having non-phosphorylated substrates. Indeed, the aldolases characterised to date form a group of remarkable sequence diversity. The most similar aldolase in terms of reaction performed is the 2-keto-3-deoxy-phosphogluconate aldolase, which catalyses the equivalent step in the classical, phosphorylated, Entner-Doudoroff pathway (Entner and Doudoroff, 1952; Kersters and De Ley, 1968). This enzyme has been successfully purified using dye ligand chromatography, with one matrix as a positive step and another matrix as a negative step (Scopes, 1984). Unfortunately, these techniques relied heavily on the nature of the enzyme to have phosphorylated substrates and its ability to be eluted by them. In the case of the KDG-aldolase therefore, this strategy was impossible. Likewise, it was not attempted to mirror a purification protocol utilised for any other aldolase, but rather to screen a large number of techniques with the aim of combining the most profitable.

It should be apparent from the results shown that purification of KDG-aldolase enzyme to homogeneity in any great quantity was not found to be possible by combining any of the methods attempted. Although successful in its end conclusion of having purified enough enzyme to obtain an N-terminal sequence, the particular method of purification

described here is clearly far from ideal, and if further protein were to be required various modifications to the protocol would be investigated.

The heat precipitation step does not increase the purity enough to warrant its use in this case as a valuable step for protein purification. The fold purification could possibly be increased by incubation for a longer period at 90°C, but this may cause the KDG-aldolase itself to precipitate, or for it to aggregate with precipitated proteins thus coming out of solution with the insoluble fraction. Use of a heat step for enzyme purification is normally reserved for the convenient separation of thermostable recombinant enzyme from mesophilic host proteins, and therefore this step would be most profitably utilised as demonstrated in chapter 5, for the purification of recombinant KDG-aldolase from *E. coli*.

Ammonium sulphate precipitation was successful in that the enzyme was purified 9 fold. However, it is inconvenient to be working with large volumes of dilute protein containing ammonium sulphate and therefore, in future purification protocols, it would be preferable to do a second cut at a higher concentration of ammonium sulphate which would precipitate the KDG-aldolase. This would serve both to separate the enzyme from those proteins that do not precipitate at that higher ammonium sulphate concentration, and to allow the resuspension of the precipitated protein in a smaller volume.

Anion exchange further purified the protein by nearly 4 fold, but not to homogeneity. As the subunit $_{Mr}$ weight was as yet undetermined, it was not possible to identify the KDG-aldolase, and therefore further purification was necessary. The method utilised, a variation on a technique used by De Montigny and Sygusch (1996), allowed identification of KDG-aldolase as either a monomeric, or homomeric, protein of approximately 32 kDa, and was able to provide enough protein to be N-terminally sequenced.

The N-terminal sequence was found to have reasonable identity to both a fructose bisphosphate aldolase and *N*-acetylneuraminate lyase subunit. The fructose bisphosphate aldolase is a monomeric 33 kDa protein that catalyses the 4th step in glycolysis of fructose 1,6-bisphosphate to dihydroxyacetone-phosphate and glyceraldehyde 3-phosphate (Witke and

Götz, 1993). Perhaps more significant is the match with the *N*-acetylneuraminate lyase subunit, also known as *N*-acetylneuraminic acid aldolase. This well characterised enzyme (Comb and Roseman, 1960; Ohta et al., 1985) catalyses the reaction between *N*-acetyl-D-mannosamine and pyruvate to *N*-acetylneuraminate, usually in this catabolic direction, in the production of sialic acids. Neither substrates, nor products are phosphorylated, and the subunit M_r of the enzyme is 32.5 kDa, similar to that found by SDS-PAGE analysis of the KDG-aldolase. Continuing comparison between the two enzymes as more information is revealed about the KDG-aldolase will be of interest.

5 CLONING, SEQUENCING AND ANALYSIS OF KDG-ALDOLASE GENE

5.1 INTRODUCTION

Having isolated the KDG-aldolase enzyme from *S. solfataricus* and obtained a 38 amino acid N-terminal sequence (chapter 4), it is possible to clone and sequence the gene from genomic DNA. This chapter describes the cloning procedure, gives the determined sequence of KDG-aldolase from *S. solfataricus* and provides an analysis of the gene sequence.

5.2 MATERIALS AND METHODS

5.2.1 Materials

Samples of genomic DNA from *Sulfolobus solfataricus* P1 (DSM 1616) and restricted genomic DNA ligated into λEMBL3 arms were prepared and kindly donated by Dr. H. Connaris (Yeats et al., 1982; Connaris et al. 1998). *E. coli* XL-1 MRA Blue cells (P2) were suppplied by Promega, Southampton, UK.

pGEM®-T Vector System and T4 DNA ligase were supplied by Promega, Madison, USA. Hybond™-N^+ membranes and blocking agent are supplied by Amersham, Little Chalfont, Bucks., UK. The High Prime® DNA labelling kit was from Boehringer Mannheim Ltd., Lewes, East Sussex. UK. Dextran sulphate was supplied by United States Biochemical, Cleveland, OH, USA. The QIAGEN Lambda Kit for maxi preparations was supplied by QIAGEN GmBH., Hilden, Germany.

The rotating hybridisation oven was the model Biometra OV4, supplied by Biometra, Göttingen, Germany. The Hyperprocessor for developing autoradiographs was from Amersham, Little Chalfont, Bucks., UK.

20 × SSC is composed of 17.53 % (w/v) NaCl, 8.82 % (w/v) sodium citrate, pH 7. SM buffer contains 0.58 % (w/v) NaCl, 0.2 % (w/v) $MgSO_4 \cdot H_2O$, 5 % (v/v) 1 M Tris-HCl (pH 7.5), and 0.02 % (w/v) gelatine solution.

Computer-based sequence analyses were performed using the GCG package (Genetics Computer Group, Inc., Madison, WI, USA).

5.2.2 PCR amplification of 114 bp fragment from genomic DNA

Amplification of genomic DNA was carried out using primers based on the 38 amino acid N-terminal sequence obtained. Oligonucleotide design was carried out using codon bias as assessed by De Vendittis and Bocchini (1996) which is shown in Table 5.1.

5' primer 5' CCWGARATWATWACWCCWAT 3'

3' primer 5' RAATAAYTTATCWATWCCYTT 3'

Further PCR amplification details are given in section 2.4.1 and the legend to Table 5.1 specifies the degenerate codes.

5.2.3 Ligation of the gDNA PCR product into pGEM®-T Vector

The gDNA PCR product was ligated into the pGEM®-T Vector System according to the manufacturer's instructions. Vector:insert ratios of 1:1, 3:1 and 6:1 were incubated for 3 h at 15°C with T4 DNA ligase in the buffer supplied.

5.2.4 Transformation of *E. coli* JM109 Competent cells

Transformations were carried out as described in section 2.4.5 with the modification that LB plates supplemented with 100 μg ampicillin/ml were also treated with X-gal (50 μl of 20 mg/ml) and IPTG (50 μl of 24 mg/ml) to facilitate blue/white screening of colonies that contain plasmid with and without insert, respectively. Plates were incubated overnight at 37°C.

5.2.5 gDNA λ library plate lysate preparation

5 ml of LB medium supplemented with 0.2% maltose and 10 mM $MgSO_4$ was inoculated with *E. coli* XL-1 MRA Blue cells (P2) and grown at 37°C with shaking to an OD_{600nm} of 1. The cells were then pelleted at 1,500 g for 10 min and the pellet resuspended in chilled 10 mM $MgSO_4$ to give a final OD_{600nm} of 0.5. The cells were stored on ice until further required (up to 24 h).

200 μl cells, in 15 ml sterile tubes, were infected with either 1 μl or 10 μl of *S. solfataricus* gDNA λ library and incubated for 20 min at 37°C. Following incubation, 3 ml

Amino acid	Codon	Frequency	Preferred codon (anti-codon)
Arg	CGA	12	AGA (TCT)
	CGC	9	
	CGG	1	
	CGT	15	
	AGA	295	
	AGG	161	
Leu	CTA	168	TTA (AAT)
	CTC	38	
	CTG	39	
	CTT	103	
	TTA	363	
	TTG	106	
Ser	TCA	127	WSW (WSW)
	TCC	57	
	TCG	34	
	TCT	110	
	AGC	63	
	AGT	134	
Thr	ACA	125	ACW (TGW)
	ACC	60	
	ACG	39	
	ACT	189	
Pro	CCA	194	CCW (GGW)
	CCC	58	
	CCG	28	
	CCT	118	
Ala	GCA	218	GCW (CGW)
	GCC	74	
	GCG	70	
	GCT	222	
Gly	GGA	242	GGW (CCW)
	GGC	55	
	GGG	81	
	GGT	217	
Val	GTA	291	GTA (CAT)
	GTC	70	
	GTG	105	
	GTT	257	
Ile	ATA	400	ATW (TAW)
	ATC	81	
	ATT	217	
Lys	AAA	337	AAR (TTY)
	AAG	327	
Asn	AAC	148	AAY (TTR)
	AAT	275	
Gln	CAA	143	CAA (GTT)
	CAG	42	
His	CAC	57	CAY (GTR)
	CAT	91	
Glu	GAA	366	GAR (CTY)
	GAG	270	
Asp	GAC	134	GAT (CTA)
	GAT	346	
Tyr	TAC	158	TAT (ATA)
	TAT	232	
Cys	TGC	7	TGT (ACA)
	TGT	27	
Phe	TTC	148	TTY (AAR)
	TTT	189	
Met	ATG	185	ATG (TAC)
Trp	TGG	97	TGG (ACC)
Start	ATG	17	ATG (TAC)
	GTG	4	
	TTG	2	
Stop	TAA	17	TAA (ATT)
	TAG	3	
	TGA	3	
Total		8871	

Table 5.1 Codon usage in *Sulfolobus solfataricus*. Frequency is calculated based on analysis of 23 protein encoding genes of different size and unrelated function, with a combined total amino acid composition of 8871. The degenerate nucleotide IUB codes used are R (A/G), W (A/T), S (C/G), Y(C/T), B (C/G/T) and N (A/C/G/T). The codon usage frequency was reproduced from De Vendittis and Bocchini (1996).

of molten LB top agarose[1] (LB broth supplemented with 0.7 % (w/v) agarose) was added to the incubating cells, mixed by inversion of the tubes and immediately plated onto pre-warmed LB plates. Once set, plates were inverted and incubated overnight at 37°C. Plates should not show confluent lysis, but isolated plaques.

5.2.6 Transfer and immobilisation of λ plaques from LB top agarose plates onto nitrocellulose filters

Hybond-N^+ membranes were laid over LB top agarose plates with λ plaques one at a time, with up to five membranes having λ plaques transferred from each plate. The membranes and plates were marked with needle holes in such a way that their corresponding orientations could be matched again at a later stage. Each membrane was left on the plate for 30 – 60 s after which time it was peeled off the plate and transferred, DNA side facing up, to a dish containing Whatman 3MM paper soaked in denaturing solution (0.5 M NaOH, 1.5 M NaCl) where it remained for 1 – 5 min. The membrane was then transferred to a second dish again containing Whatman 3MM paper soaked in neutralising solution (1.5 M NaCl, 0.5 M Tris-HCl, pH 7.4) for 5 min. Filters were then rinsed in 2 × SSC, and placed to dry, DNA side up, on dry Whatman paper. Finally, filters were baked at 80°C for 30 min to fix the DNA.

5.2.7 [α^{32}P]dCTP labelling of 114 bp DNA probe using High Prime®

The High Prime® "random primed" DNA labelling kit method is based on hybridisation of oligonucleotides of all possible sequences to the denatured DNA fragment to be labelled (Feinberg and Vogelstein, 1983; 1984). Klenow polymerase uses the 3'OH termini of the random oligonucleotides as primers with which to synthesise the complementary DNA strand, incorporating [α^{32}P]dCTP into the newly synthesised strand.

[1] LB top agarose is kept molten at 48°C.

5.2.8 Hybridisation of probe to nitrocellulose filters

Each Hybond™-N$^+$ membrane with immobilised DNA was treated with 5 ml pre-hybridisation solution to which 50 µl of freshly sheared salmon testes DNA was added to prevent non-specific binding of the probe to the membrane. Pre-hybridisation solution contains 5 % (w/v) dextran sulphate, 0.5 % (w/v) blocking agent, 0.1 % (w/v) SDS in 5× SSC. Membranes were incubated at 55°C for 2 h in a rotating hybridisation oven.

Following pre-hybridisation, 2 µl per membrane of freshly denatured labelled probe was added to each tube, and further incubated at 55°C overnight.

Following hybridisation, the hybridisation solution was removed, and the filters washed twice in 5 ml (per filter) 1 × SSC, 0.1 % (w/v) SDS for 20 min at 55°C, and then twice in 0.5 × SSC, 0.1 % (w/v) SDS for 20 min at 55°C.

5.2.9 Autoradiograph development

Washed membranes from section 5.2.8 were wrapped in cling film and placed in an X-ray cassette, overlayed with X-ray film. The cassette was stored at -70°C for up to 24 h. The autoradiographs were developed using a Hyperprocessor

5.2.10 Removal of positive λ clone plaques from screened plates

Having identified positive λ clones on autoradiographic film, the corresponding plaque is identified on the TOP agar LB plate. The plaque can be removed as a 'plug eluate' by stabbing the plaque with a sterile Pasteur pipette, and expelling the 'plug' into 100 µl SM buffer supplemented with 10 µl chloroform to prevent bacterial growth. 'Plug eluate' samples should be gently mixed by inversion at 4°C for 3-4 h, before storage at that temperature.

5.2.11 λ DNA maxiprep

Preparation of the positive clone of λ DNA for sequencing of the KDG-aldolase gene was carried out using the QIAGEN Lambda Kit for maxi preparations. Prior to following the manufacturer's instructions a liquid culture of infected XL1 Blue MRA cells (P2) should be

grown. 500 µl of an overnight culture of *E. coli* XL-1 MRA Blue cells (P2) in LB medium was added to 10 µl of a plug eluate sample from the previous section. This was incubated for 20 min at 37°C, and then used to infect 200 ml LB medium supplemented with 0.2 % (w/v) maltose, 10 mM $MgSO_4$. This culture was incubated overnight at 37°C, and then for a further 15 min with 40 µl of chloroform added. Cellular debris was removed by centrifugation at 6.5 g for 10 min, and the supernatant retained. The QIAGEN protocol was followed from this point.

Procedures used in this chapter that are not mentioned in this section will be found in chapter 2.

5.3 RESULTS

5.3.1 Cloning the KDG-aldolase gene from *S. solfataricus* genomic DNA

The cloning and sequencing procedure used for obtaining the KDG-aldolase gene is summarised diagrammatically in Figure 5.1.

5.3.1.1 Amplification of 114 bp fragment from genomic DNA by PCR

Using the 38 amino acid N-terminal sequence in conjunction with the codon usage table (Table 5.1), degenerate oligonucleotide PCR primers were designed.

5' primer 5' CCWGARATWATWACWCCWAT 3'

3' primer 5' RAATAAYTTATCWATWCCYTT 3'

PCR of *S. solfataricus* genomic DNA gave fragments of the expected size of 114 bp (Figure 5.2). Negative controls omitting either one of the primers, or the template gDNA, all correctly yielded no product.

5.3.1.2 Sequencing of the 114 bp fragment

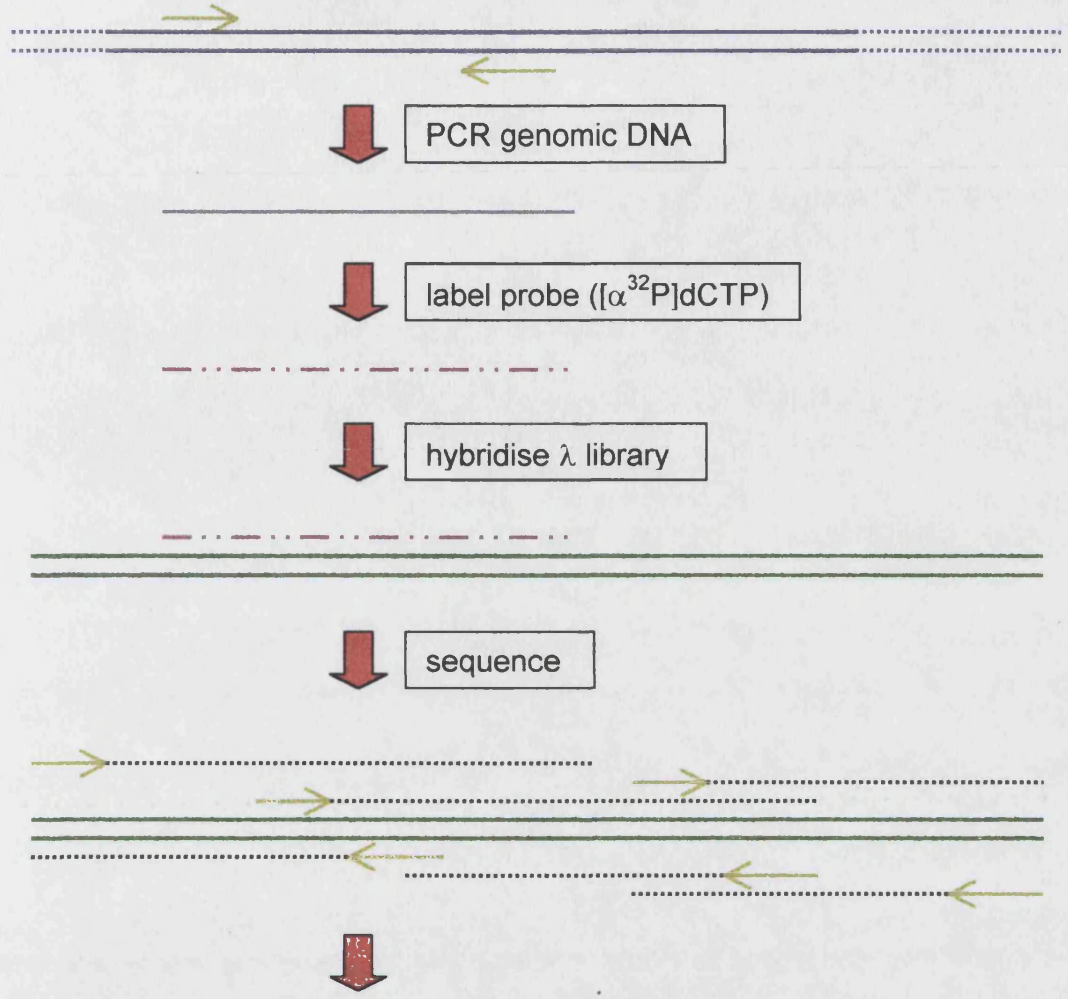
The PCR generated 114 bp fragment was prepared for sequencing by sub-cloning into pGEM-T, and was then sequenced using universal primers (Figure 5.3).

5.3.1.3 Primary lambda library screen

The 114 bp fragment was labelled and used to screen immobilised lambda DNA. The developed autoradiograph from the first lamda library screen indicated a large number of positive plaques on the corresponding TOP agar LB plates. Several 'plug eluates' were removed and λ DNA prepared from them. A second set of plate lysates were created using this DNA, and screened as for the primary screen.

5.3.1.4 Secondary lambda library screen

The second developed autoradiograph again indicated a large number of positive plaques on the corresponding TOP agar LB plates. A number of 'plug eluates' were removed




```
ATGCCAGAAATCATAACTCCAATCATAACCCCATTCACTAAAGATAATAGAATAGATAAGGAAAAATTAAAGATACAT
M  P  E  I  I  T  P  I  I  T  P  F  T  K  D  N  R  I  D  K  E  K  L  K  I  H
GCGGAGAATCTCATTAGGAAGGGAATAGATAAGTTGTTCGTCAACGGTACTACTGGTCTTGGTCCTTCGTTATCTCCA
A  E  N  L  I  R  K  G  I  D  K  L  F  V  N  G  T  T  G  L  G  P  S  L  S  P
GAGGAGAAGTTAGAGAACTTAAAGGCAGTTTATGACGTCACCAATAAGATAATATTTCAAGTTGGTGGATTGAATCTA
E  E  K  L  E  N  L  K  A  V  Y  D  V  T  N  K  I  I  F  Q  V  G  G  L  N  L
GACGATGCTATAAGATTGGCTAAATTAAGTAAAGACTTTGATATTGTCGGTATAGCCTCGTATGCTCCATATTATTAC
D  D  A  I  R  L  A  K  L  S  K  D  F  D  I  V  G  I  A  S  Y  A  P  Y  Y  Y
CCAAGAATGTCTGAGAAGCATTTGGTAAAGTATTTTAAGACCTTGTGTGAAGTATCTCCACACCCTGTCTATTTGTAC
P  R  M  S  E  K  H  L  V  K  Y  F  K  T  L  C  E  V  S  P  H  P  V  Y  L  Y
AATTACCCGACGGCAACGGGAAAAGACATAGATGCAAAAGTCGCTAAAGAGATAGGCTGTTTTACTGGAGTAAAGGAT
N  Y  P  T  A  T  G  K  D  I  D  A  K  V  A  K  E  I  G  C  F  T  G  V  K  D
ACTATTGAAAACATAATTCACACCTTAGACTACAAACGTCTAAATCCTAACATGTTAGTATATAGTGGCTCTGATATG
T  I  E  N  I  I  H  T  L  D  Y  K  R  L  N  P  N  M  L  V  Y  S  G  S  D  M
TTAATAGCAACGGTAGCTTCTACGGGTTTAGATGGTAATGTTGCAGCAGGTTCGAATTATCTTCCAGAGGTTACTGTG
L  I  A  T  V  A  S  T  G  L  D  G  N  V  A  A  G  S  N  Y  L  P  E  V  T  V
ACAATTAAGAAATTGGCTATGGAAAGGAAAATTGATGAAGCACTTAAGTTACAATTCCTTCATGACGAGGTAATAGAG
T  I  K  K  L  A  M  E  R  K  I  D  E  A  L  K  L  Q  F  L  H  D  E  V  I  E
GCGTCTAGAATATTTGGGAGCTTATCTTCAAATTACGTATTAACCAAGTATTTCCAAGGATACGATTTAGGATATCCT
A  S  R  I  F  G  S  L  S  S  N  Y  V  L  T  K  Y  F  Q  G  Y  D  L  G  Y  P
AGACCTCCAATATTCCCACTAGATGATGAAGAAGAAAGGCAGCTAATTAAGAAAGTTGAGGGTATAAGGGCGAAACTT
R  P  P  I  F  P  L  D  D  E  E  E  R  Q  L  I  K  K  V  E  G  I  R  A  K  L
GTAGAGCTTAAAATATTGAAAGAATAG
V  E  L  K  I  L  K  E  *
```

Figure 5.1 Schematic diagram of procedure for cloning and sequencing KDG-aldolase.

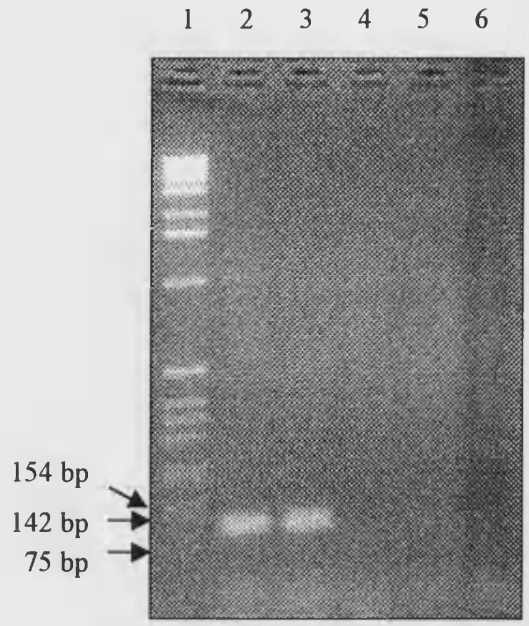


Figure 5.2 Agarose electrophoresis identification of PCR product. The gel shows the PCR product (lanes 2 and 3) to be approximately 114 bp in comparison to DNA size markers (lane 1). Remaining lanes are lane 4, no 5' primer; lane 5, no 3' primer; and lane 6, no template gDNA.

```
   CCTGAGATTATTACTCCTATCATAACCCCATTCACTAAAGATAATAGAATAGATAAG
1  ---------+---------+---------+---------+---------+-------

   GGACTCTAATAATGAGGATAGTATTGGGGTAAGTGATTTCTATTATCTTATCTATTC

   P  E  I  I  T  P  I  I  T  P  F  T  K  D  N  R  I  D  K

   GAAAAATTAAAGATACATGCGGAGAATCTCATTAGGAAGGGAATTGATAAATTATTC
60 --+---------+---------+---------+---------+---------+----

   CTTTTTAATTTCTATGTACGCCTCTTAGAGTAATCCTTCCCTTAACTATTTAATAAG

   E  K  L  K  I  H  A  E  N  L  I  R  K  G  I  D  K  L  F
```

Figure 5.3 Sequence of 114 bp PCR product with amino acid translation. Forward and reverse DNA sequences were matched and translated using GCG. Bases indicated in bold type indicate the regions amplified with degenerate primers. The translated amino acid sequence is identical to the 38 amino acid N-terminal sequence obtained in chapter 4.

and λ DNA prepared from them. Using the primers created for amplification of the 114 bp fragment (section 5.2.3) 'true positives' that contained this region of DNA were identified by PCR. One of these 'true positives' was selected for further DNA sequencing.

5.3.2 The KDG-aldolase gene sequence from *S. solfataricus*

The entire KDG-aldolase gene and flanking regions were sequenced in both directions with suitable overlaps to eliminate all ambiguities. The complete nucleotide sequence and deduced amino acid sequence of the *S. solfataricus* KDG aldolase gene is shown in Figure 5.4. The GenBank database accession number is AJ224174.

5.3.3 Analysis of the KDG-aldolase gene sequence from *S. solfataricus*

5.3.3.1 Primary sequence analysis

The KDG-aldolase gene open reading frame was determined at the 5' end by complete agreement of the translated sequence with the 38 amino acid N-terminal sequence (section 5.3.8). The 3' end was determined by translation to give a stop codon (TAG).

The open reading frame sequence translates to give a protein 293 amino acids long, with a polypeptide molecular weight of 32976 calculated from its primary sequence. This value is in agreement with that of 32 kDa determined by SDS-PAGE (section 4.3.7). The isoelectric point of the protein is 7.05, which may account for its weak binding at pH 8.5 in anion exchange chromatography in section 4.3.5. The calculated extinction coefficient (280 nm) is 20,600 $M^{-1}cm^{-1}$. These results were determined using the GCG package.

5.3.3.2 Identity searching and sequence alignments

Database searches and sequence alignments were analysed using Basic Local Alignment Search Tool, version 2.0.9 (BLAST, Altschul et al., 1990; Altschul et al., 1997). High BLAST scores were observed with *N*-acetylneuraminate lyase from *Haemophilus influenzae* (27 % identity) and *E. coli* (24 %), and dihydrodipicolinate synthase from *Aquifex aeolicus* (27 %), *Methanococcus janaschii* (26 %), and *Bacillus subtilis* (26 % identity).

```
-90                -70                -50                -30                -10
TAATAGAACAGCTAAGAGCTGAACCAATACCATTAGATGTAATTGAAGAACCGGTTTGGGTCGTCAAGGGAACCTGGAAGAATTATGGTGTTTGAGG
         10                 30                 50                 70                 90
ATGCCAGAAATCATAACTCCAATCATAACCCCATTCACTAAAGATAATAGAATAGATAAGGAAAAATTAAAGATACATGCGGAGAATCTCATTAGGA
M  P  E  I  I  T  P  I  I  T  P  F  T  K  D  N  R  I  D  K  E  K  L  K  I  H  A  E  N  L  I  R  K
          110                130                150                170                190
AGGGAATAGATAAGTTGTTCGTCAACGGTACTACTGGTCTTGGTCCTTCGTTATCTCCAGAGGAGAAGTTAGAGAACTTAAAGGCAGTTTATGACGT
  G  I  D  K  L  F  V  N  G  T  T  G  L  G  P  S  L  S  P  E  E  K  L  E  N  L  K  A  V  Y  D  V
             210                230                250                270                290
CACCAATAAGATAATATTTCAAGTTGGTGGATTGAATCTAGACGATGCTATAAGATTGGCTAAATTAAGTAAAGACTTTGATATTGTCGGTATAGCC
 T  N  K  I  I  F  Q  V  G  G  L  N  L  D  D  A  I  R  L  A  K  L  S  K  D  F  D  I  V  G  I  A
                310                330                350                370
TCGTATGCTCCATATTATTACCCAAGAATGTCTGAGAAGCATTTGGTAAAGTATTTTAAGACCTTGTGTGAAGTATCTCCACACCCTGTCTATTTGT
S  Y  A  P  Y  Y  Y  P  R  M  S  E  K  H  L  V  K  Y  F  K  T  L  C  E  V  S  P  H  P  V  Y  L  Y
390                410                430                450                470
ACAATTACCCGACGGCAACGGGAAAAGACATAGATGCAAAAGTCGCTAAAGAGATAGGCTGTTTTACTGGAGTAAAGGATACTATTGAAAACATAAT
  N  Y  P  T  A  T  G  K  D  I  D  A  K  V  A  K  E  I  G  C  F  T  G  V  K  D  T  I  E  N  I  I
   490                510                530                550                570
TCACACCTTAGACTACAAACGTCTAAATCCTAACATGTTAGTATATAGTGGCTCTGATATGTTAATAGCAACGGTAGCTTCTACGGGTTTAGATGGT
 H  T  L  D  Y  K  R  L  N  P  N  M  L  V  Y  S  G  S  D  M  L  I  A  T  V  A  S  T  G  L  D  G
      590                610                630                650                670
AATGTTGCAGCAGGTTCGAATTATCTTCCAGAGGTTACTGTGACAATTAAGAAATTGGCTATGGAAAGGAAAATTGATGAAGCACTTAAGTTACAAT
N  V  A  A  G  S  N  Y  L  P  E  V  T  V  T  I  K  K  L  A  M  E  R  K  I  D  E  A  L  K  L  Q  F
         690                710                730                750                770
TCCTTCATGACGAGGTAATAGAGGCGTCTAGAATATTTGGGAGCTTATCTTCAAATTACGTATTAACCAAGTATTTCCAAGGATACGATTTAGGATA
  L  H  D  E  V  I  E  A  S  R  I  F  G  S  L  S  S  N  Y  V  L  T  K  Y  F  Q  G  Y  D  L  G  Y
            790                810                830                850                870
TCCTAGACCTCCAATATTCCCACTAGATGATGAAGAAGAAAGGCAGCTAATTAAGAAAGTTGAGGGTATAAGGGCGAAACTTGTAGAGCTTAAAATA
 P  R  P  P  I  F  P  L  D  D  E  E  E  R  Q  L  I  K  K  V  E  G  I  R  A  K  L  V  E  L  K  I
               890                910                930                950                970
TTGAAAGAATAGTATACTATCATGGTTGATGTAATAGCTTTGGGAGAGCCTTTAATCCAATTTAACTCTTTTAACCCTGGTCCGTTGAGATTCGTAA
L  K  E  *
                  990                1010               1030               1050
ACTATTTTGAAAAACATGTAGCAGGATCTGAGTTAAATTTCTGCATTGCTGTTGTTAGGAATCATTTATCATGTAGTTTAATAGCAAGAGTAGGGAA

1070               1090               1110               1130
TGATGAAGTTTGGTAAGAACATTATAGAATATTCTAGAGCTCAAGGTATTGATACTAGCCATATAAAGGTTGATAA
```

Figure 5.4 Complete nucleotide sequence of the KDG-aldolase gene from *S. solfataricus* and its flanking regions. The gene is numbered 1 - 882, and the corresponding one letter amino acid codes are indicated below.

Significant identities were also seen with the *Rhizobium meliloti* MosA protein (23 %) and the *Bacillus subtilis* probable 5-dehydro-4-deoxyglucarate dehydratase (24 %). Of unknown relevance were matches (26 % identity) with the hypothetical proteins from the *E. coli* genome, a hypothetical 34.8 kDa protein in feci-fimb intergenic region and a 33.3 kDa protein in the perr-argf intergenic region. These results are summarised in Table 5.2.

Accession number	Protein name	Organism	BLAST value	Identity (%)
P75682	Hypothetical 33.3 kD protein in the perr-argf intergenic region	*Escherichia coli*	4e-21	26
P39359	Hypothetical 34.8 kDa protein in feci-fimb intergenic region	*Escherichia coli*	1e-20	26
P06995	*N*-acetylneuraminate lyase	*Escherichia coli*	4e-16	24
P44539	*N*-acetylneuraminate lyase	*Haemophilus influenzae*	8e-20	27
Q04796	Dihydrodipicolinate synthase	*Bacillus subtilis*	1e-15	26
Q57695	Dihydrodipicolinate synthase	*Methanococcus jannashii*	2e-15	26
AE000725	Dihydrodipicolinate synthase	*Aquifex aeolicus*	1e-13	27
P43797	Dihydrodipicolinate synthase	*Haemophilus influenzae*	2e-13	25
P05640	Dihydrodipicolinate synthase	*Escherichia coli*	6e-09	25
Q07607	MosA protein	*Rhizobium meliloti*	8e-09	23
P42235	5-Keto-4-deoxy-glucarate dehydratase	*Bacillus subtilis*	5e-05	24

Table 5.2 Table of selected results from BLAST search using KDG-aldolase as the target sequence (BLAST, Altschul et al., 1990; Altschul et al., 1997). GenBank CDS translations, PDB, SwissProt, PIR and PRF databases were searched.

Sequence alignments were produced using CLUSTAL (Higgins and Sharp, 1988). An alignment of selected sequences, coloured to indicate regions of high identity and similarity

across all the sequences, is given in Figure 5.5. Figure 5.6 indicates the regions in the aligned sequences that show identity or similarity *specifically* to KDG-aldolase.

5.3.3.3 Secondary structure prediction

Figure 5.7 illustrates the assignment of secondary structure as predicted by the method of Sequence Annotated by Structure (SAS; Milburn et al., 1998). It is noted that KDG-aldolase is predicted to have an eight stranded α/β-barrel configuration similar to that of both NAL and DHDPS, for which the three dimensional structures have been solved (Izard et al., 1994; Mirwaldt et al., 1995).

5.3.4 Analyses of upstream and downstream regions

The sequenced regions upstream and downstream to the KDG-aldolase gene were analysed, primarily with the aim of identifying a Shine-Dalgarno sequence and promoter site upstream, and a transcriptional stop-signal (poly dT tract) downstream. None of these were identified. However, the upstream sequence (97 bp) translates into a potential ORF that terminates 2 bp before the KDG-aldolase start codon. No matches to this partial sequence were found in the databases. A second potential ORF (180 bp), begins 9 bp downstream of the KDG-aldolase gene, and this partial translated sequence shows the highest identity score with residues 8 - 63 of a probable fructokinase from *Pyrococcus horikoshii* (41 %; Acc. No. AP000006). Also of importance is the 30 % identity with residues 3 – 62 of KDG kinase from *Bacillus subtilis* (Acc. No. P50845). The absence of initiation and termination elements and the translation of the flanking regions suggest that the KDG-aldolase gene is part of an operon.

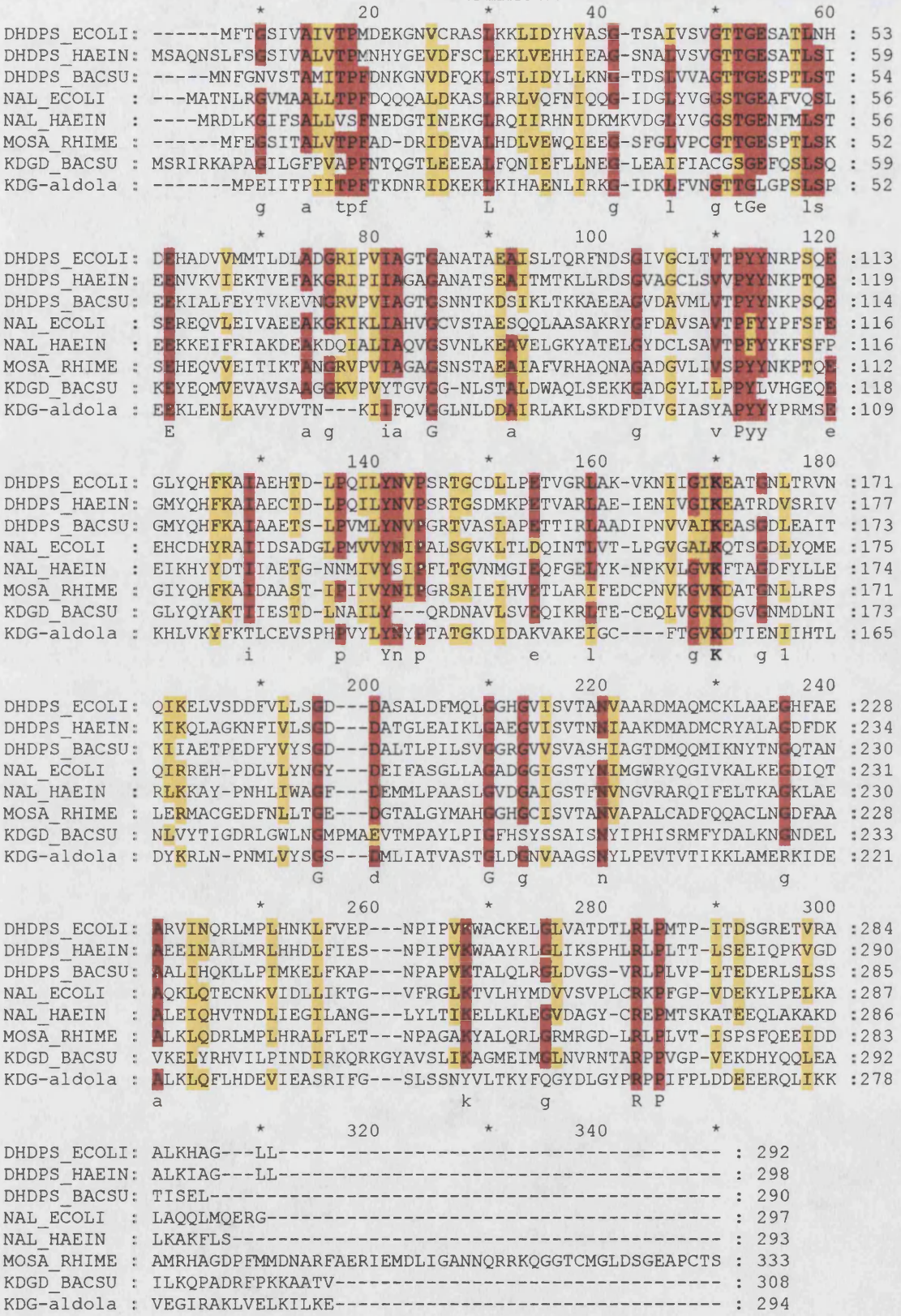


Figure 5.5 Sequence alignment of DHDPS from *E. coli, H. influenzae and E. coli, H. influenzae*, NAL from *E. coli* and *H. influenzae*, MosA protein from *R. meliloti*, KDGD from B. subtilis and KDG-aldolase from *S. solfataricus*. Six or more identical residues that align are highlighted in red, those that are similar are highlighted in yellow. The alignment was produced by CLUSTAL (Higgins and Sharp, 1988).

Figure 5.6 Sequence alignment of DHDPS from *E. coli, H. influenzae* and *E. coli, H. influenzae*, NAL from *E. coli* and *H. influenzae, MosA* protein from *R. meliloti*, KDGD from *B. subtilis* and KDG-aldolase from *S. solfataricus*. Aligned residues identical to those in KDG-aldolase are shown in blue, those that are similar are shown in turquoise. The alignment was produced by CLUSTAL (Higgins and Sharp, 1988).

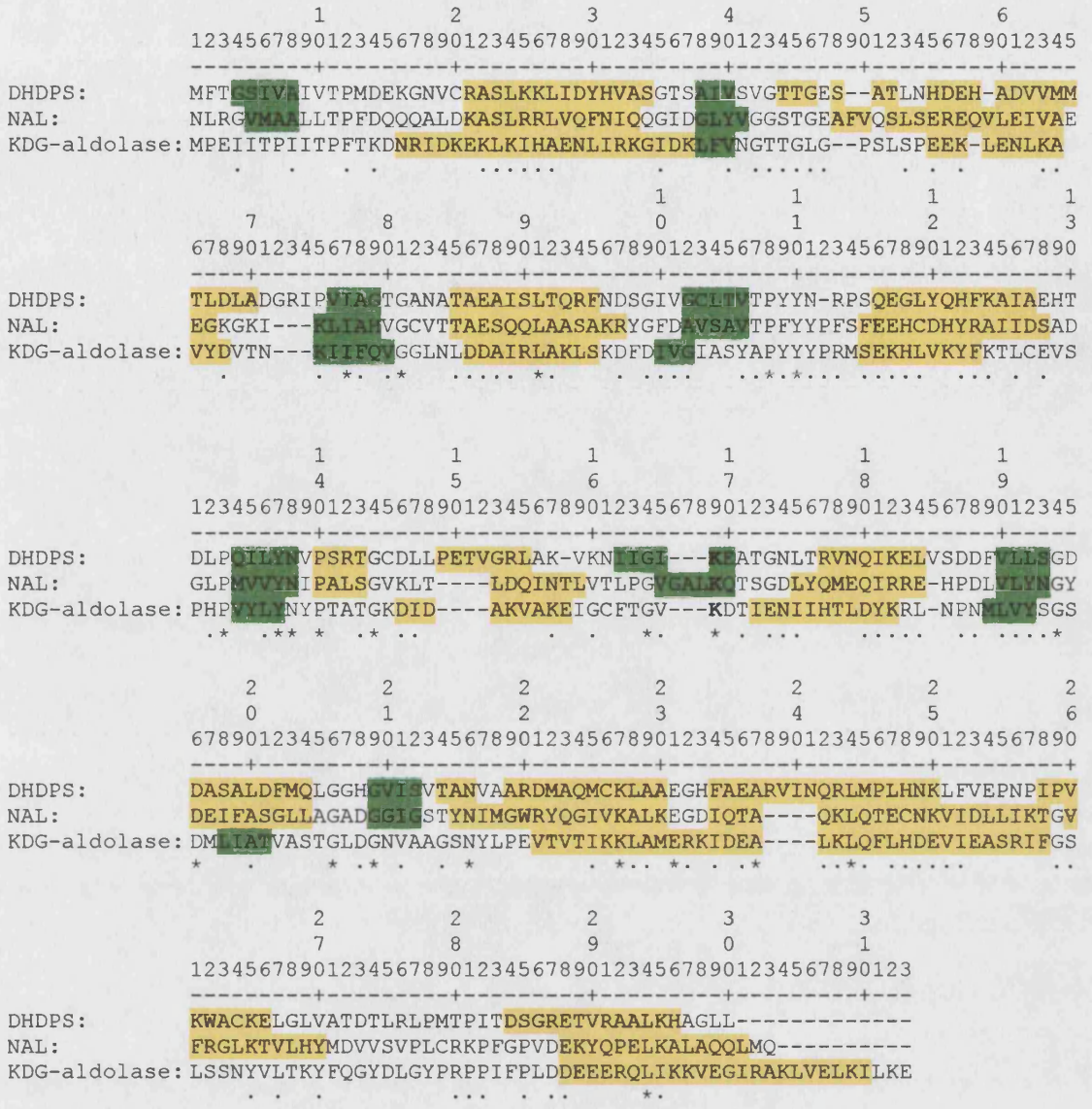


Figure 5.7 Secondary structure alignment of DHDPS and NAL from *E. coli*, and KDG-aldolase from *S. solfataricus*. α-helices are indicated by yellow highlighting and β-sheets are indicated by green highlighting. Residues identical in the three sequences are marked with (*), those identical in two are marked with (·). DHDPS and NAL secondary structure is given from known 3D structures (Mirwaldt et al., 1995; Izard et al., 1994). Secondary structure predictions for KDG-aldolase and the alignment are produced by Sequences Annotated by Structure, Version 1.5 (Milburn et al., 1998).

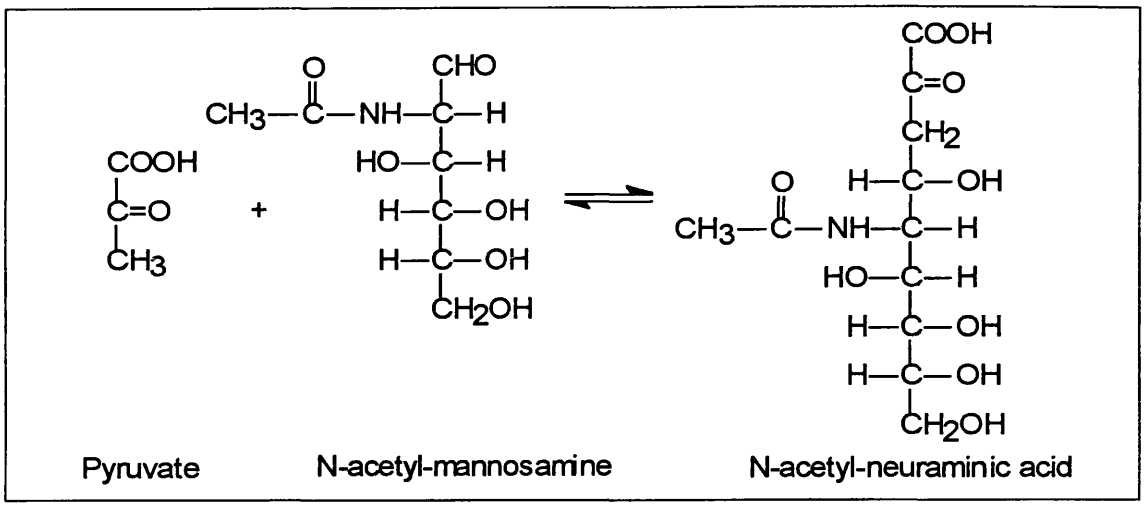
5.4 DISCUSSION

A BLAST database search with the amino acid sequence of KDG-aldolase from *S. solfataricus* shows significant identity to the enzymes NAL and DHDPS from a variety of organisms. These enzymes have been proposed to form a new sub-group of the family of $(\alpha/\beta)_8$ proteins, based on both structural and functional relationships within the new group. Also included within this novel sub-group, are the enzymes *trans-o*-hydrozybenzylidenepyruvate hydratase aldolase, D-4-deoxy-5-oxoglucarate dehydratase, and the MosA protein (Lawrence et al., 1997). D-4-deoxy-5-oxoglucarate dehydratase, also known as 2-keto-3-deoxy-glucarate dehydratase, from *B. subtilis*, and the MosA protein from *R. meliloti*, have 24 % and 23 % identity to KDG-aldolase respectively.

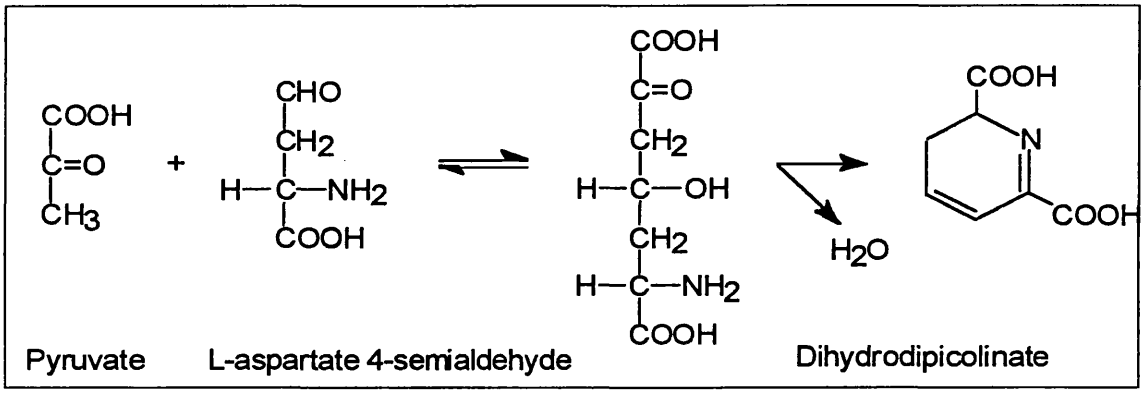
NAL is a key enzyme in sialic acid metabolism (Schauer, 1982), regulating the intracellular concentration of sialic acid to prevent toxic accumulation (Vimr and Troy, 1985a; 1985b). The enzyme has been isolated from some bacteria and from mammalian tissue (Schauer, 1982). The reaction catalysed by NAL is the reversible aldol condensation between *N*-acetyl-D-mannosamine and pyruvate to produce *N*-acetyl-D-neuraminic acid (Figure 5.8a). Particular interest in the sialic acids, and thereby NAL's ability to catalyse sialic acid production, has arisen due to their important role in a range of biomolecular recognition processes, particularly at the cell surface (Schauer, 1982).

DHDPS is the first enzyme unique to lysine biosynthesis in the diaminopimelate pathway of bacteria and higher plants (Shedlarski and Gilvarg, 1970). The enzyme catalyses the reversible condensation of pyruvate and L-aspartate-β-semialdehyde to dihydrodipicolinate (Figure 5.8b).

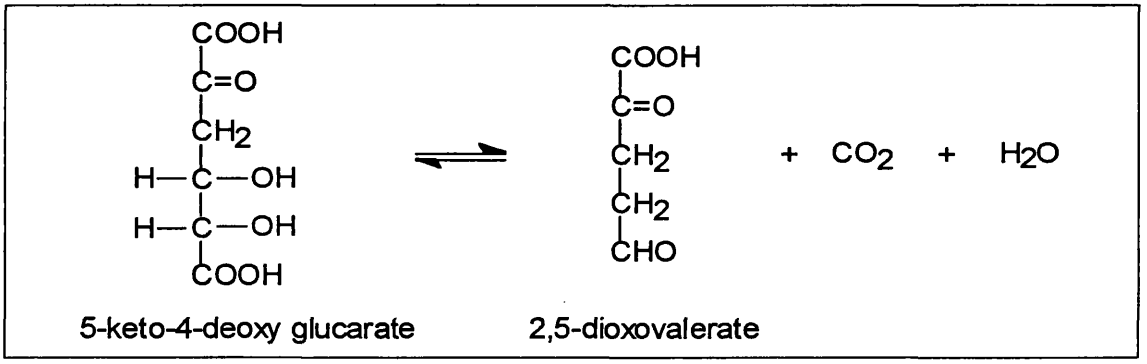
5-keto-D-4-deoxy-glucarate dehydratase or 2-keto-3-deoxy-glucarate dehydratase, is involved in the metabolism of glucarate (Jeffcoat et al., 1969a; 1969b). The reaction catalysed by this enzyme is shown in Figure 5.8c. The MosA protein is thought to catalyse



(a) *N*-acetylneuraminate lyase reaction



(b) Dihydrodipicolinate synthase reaction



(c) 2-keto-3-deoxy-glucarate dehydratase reaction

Figure 5.8 Linear form representations of the reactions catalysed by NAL (a), DHDPS (b), and KDGD (c).

the addition of a methyl group to *scylloinosamine* (Murphy et al., 1993; Rao et al., 1995) and is only a tentative member of the sub-group due to lack of clarity concerning the exact reaction and its mechanism (Lawrence et al., 1997). *Trans-o*-hydrozybenzylidenepyruvate hydratase aldolase is involved in napthalene degradation (Eaton, 1994). This enzyme did not have any significant identity with KDG-aldolase and shows average identities between 15 and 20 % with the other members of this sub-group of enzymes.

The three-dimensional structures of both NAL and DHDPS from *E. coli* have been solved by X-ray crystallography. These enzymes exist as homotetramers, with each subunit possessing eight-stranded β-barrels surrounded by alpha helices, with 222 symmetry (Figure 5.9; Izard, et al., 1994; Mirwaldt et al., 1995). Using information based on the known structures of these enzymes, in conjunction with the sequence alignments in this chapter (Figures 5.5, 5.6 and 5.7), various predictions concerning KDG-aldolase can be made. Significant identities, 24 % and 25 %, with the NAL and DHDPS enzymes from *E. coli* are given for the KDG-aldolase, and from the alignment in Figure 5.7 it is apparent that the conserved residues are clustered about the α/β barrel, with less conservation in the area of the carboxy-terminal α-helical elaboration. This, taken with the predicted secondary structure aligning well with the known helices and sheets of both NAL and DHDPS, would suggest that KDG-aldolase adopts a similar β-barrel structure.

The three-dimensional structure of KDPG is also known to be of an eight stranded β-barrel formation. However, this enzyme has an additional N-terminal α-helix compared to the NAL and DHDPS structures, is lacking structural elements characterising the NAL sub-family, exists as a trimer as opposed to a tetramer, and represents a more distant member of the $(\alpha/\beta)_8$ family (Mavridis et al., 1982; Lawrence et al., 1997). This enzyme and the MosA protein – a tentative member of the group – are not considered further in the ensuing discussion below.

Figure 5.9 Ribbon plot of the *E. coli* NAL homotetramer. Each momomer is coloured differently.

Both NAL and DHDPS catalysed aldol condensation reactions proceed via Schiff base formation between the carbonyl group of pyruvate and an ε-amino group of a lysine residue at the active site of the enzyme (Figure 5.10; Aisaka et al., 1991; Shedlarski and Gilvarg, 1970; Laber, et al., 1992). The sequence alignment, Figure 5.5, illustrates the absolute conservation of this lysine residue: KDG-aldolase, Lys155; *E.coli* NAL, Lys165; *E. coli* DHDPS, Lys161. The presence of this aligning lysine residue in KDG-aldolase would suggest that this enzyme mechanism proceeds via a similar Schiff-base formation with the pyruvate or KDG carbonyl group. Experimental data supporting this proposed mechanism are described and discussed in chapter 7.

12 core residues with side chains extending from the β-strands towards the interior of the barrel have been identified in the NAL structure, of which 6 are conserved in DHDPS (Izard, et al., 1994). Based on sequence alignments (Figure 5.11), 5 of these residues are present in KDG-aldolase (Ile77[2], Tyr137, Lys165, Tyr187 and Gly189), and two are conservatively exchanged (Tyr43 → Phe, Ile206 → Val). Izard et al. (1994) also identified in NAL a putative enzyme active site formed by a deep pocket at the carboxy-terminal end of the barrel (Figure 5.12), which is the location of the active site in all known α/β-barrel enzymes (Farber and Petsko, 1990). Of the nine amino acids that line this pocket, in KDG-aldolase 5 are conserved (Thr48, Tyr137, Lys165, Thr167 and Gly189), one is conservatively exchanged (Ser47 → Thr), and three are different (Ala11→ Pro, Ile 139 → Tyr and Tyr 190 → Ser). DHDPS has a similar conservation of these residues along a crevice of almost 30 Å, which can be accessed at the C-terminal end of the barrel via an entrance lined by Asp187 and Asp188 (Mirwaldt et al, 1995). The acidic nature of Asp 188 (191 in *E.coli* NAL) is highly conserved across the superfamily. The high degree of residue conservation suggests that the enzymes in this superfamily have a similar fold and active site environment.

[2] all amino acid numbering, unless otherwise specified, is with respect to *E. coli* NAL

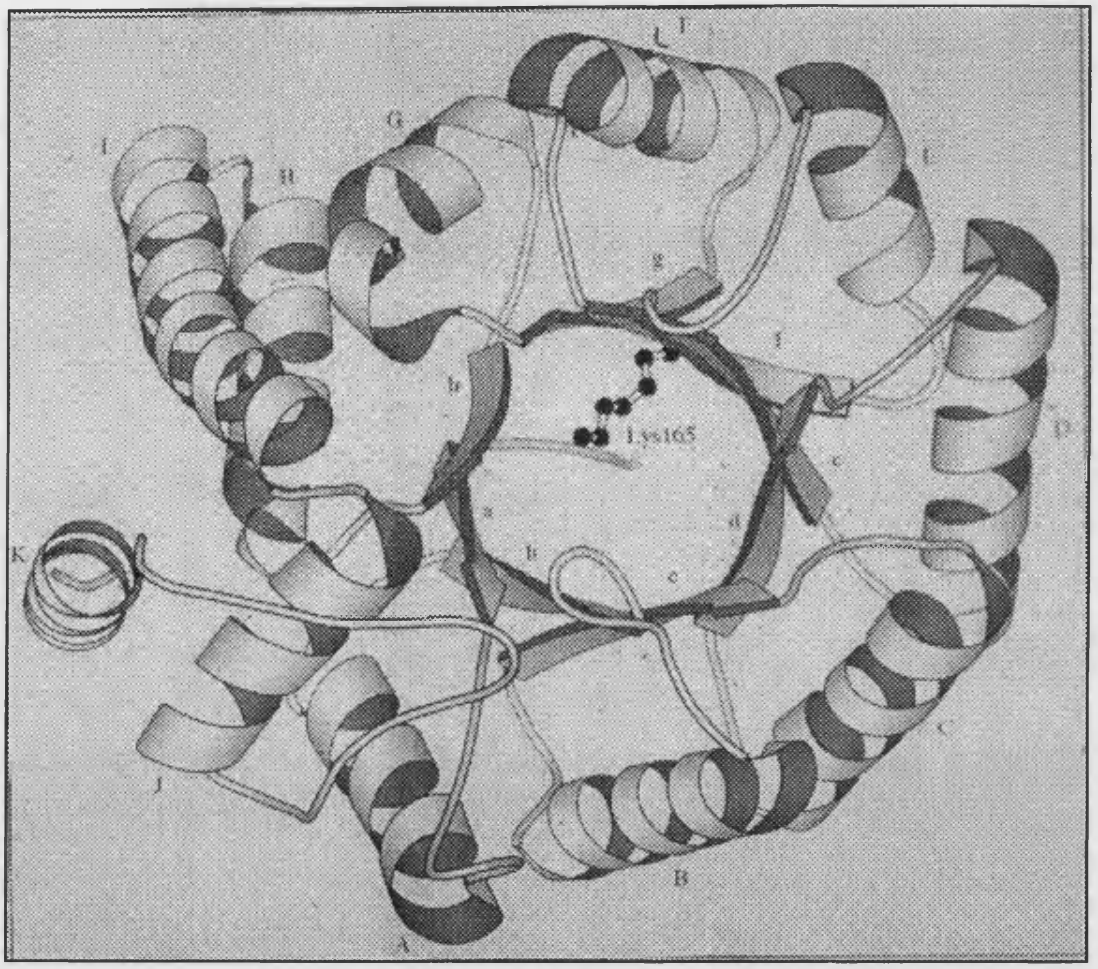


Figure 5.10 Secondary structural elements of *E. coli* NAL monomer. The monomer is viewed down the β-barrel axis from the carboxy-terminal end. The Schiff-base forming Lys165 is shown in ball and stick representation. Figure reproduced from Izard et al. (1994), where it was created using MOLSCRIPT (Kraulis et al., 1991).

```
                    *        20          *        40          *        60
DHDPS_ECOLI: ---MFTGSIVAIVTPMDEKGNVCRASLKKLIDYHVASGTSAIVSVGTTGESATLNHDEHA : 57
NAL_ECOLI  : MATNLRGVMAALLTPFDQQQALDKASLRRLVQFNIQQGIDGLYVGGSTGEAFVQSLSERE : 60
KDG-aldola : ----MPEIITPIITPFTKDNRIDKEKLKIHAENLIRKGIDKLFVNGTTGLGPSLSPEEKL : 56

                    *        80          *       100          *       120
DHDPS_ECOLI: DVVMMTLDLADGRIPVIAGTGANATAEAISLTQRFNDSGIVGCLTVTPYYNRPSQEGLYQ :117
NAL_ECOLI  : QVLEIVAEEAKGKIKLIAHVGCVSTAESQQLAASAKRYGFDAVSAVTPFYYPFSFEEHCD :120
KDG-aldola : ENLKAVYDVTN---KIIFQVGGLNLDDAIRLAKLSKDFDIVGIASYAPYYYPRMSEKHLV :113

                    *       140          *       160          *       180
DHDPS_ECOLI: HFKAIAEHTD-LPQILYNVPSRTGCDLLPETVGRLAKVKNIIGIKEATGNLTRVNQIKEL :176
NAL_ECOLI  : HYRAIIDSADGLPMVVYNIPALSGVKLTLDQINTLVTLPGVGALKQTSGDLYQMEQIRRE :180
KDG-aldola : KYFKTLCEVSPHPVYLYNYPTATGKDIDAKVAKEIGC---FTGVKDTIENIIHTLDYKRL :170

                    *       200          *       220          *       240
DHDPS_ECOLI: VSDDFVLLSGDDASALDFMQLGGHGVISVTANVAARDMAQMCKLAAEGHFAEARVINQRL :236
NAL_ECOLI  : H-PDLVLYNGYDEIFASGLLAGADGGIGSTYNIMGWRYQGIVKALKEGDIQTAQKLQTEC :239
KDG-aldola : N-PNMLVYSGSDMLIATVASTGLDGNVAAGSNYLPEVTVTIKKLAMERKIDEALKLQFLH :229

                    *       260          *       280          *       300
DHDPS_ECOLI: MPLHNKLFVEPNPIPVKWACKELGLVATDTLRLPMTP-ITDSGRETVRAALKHAG---LL :292
NAL_ECOLI  : NKVIDLLIKTGVFRGLKTVLHYMDVVSVPLCRKPFGP-VDEKYLPELKALAQQLMQERG- :297
KDG-aldola : DEVIEASRIFGSLSSNYVLTKYFQGYDLGYPRPPIFPLDDEEERQLIKKVEGIRAKLVEL :289

                    *       320          *       340
DHDPS_ECOLI: ------------------------------------- :
NAL_ECOLI  : ------------------------------------- :
KDG-aldola : KILKE---------------------------------- : 294
```

Figure 5.11 Sequence alignment of DHDPS and NAL from *E. coli*, and KDG-aldolase from *S. solfataricus*. Putative active site residues are highlighted in yellow and barrel core residues are marked in red (Izard et al., 1994). The alignment was produced by CLUSTAL (Higgins and Sharp, 1988).

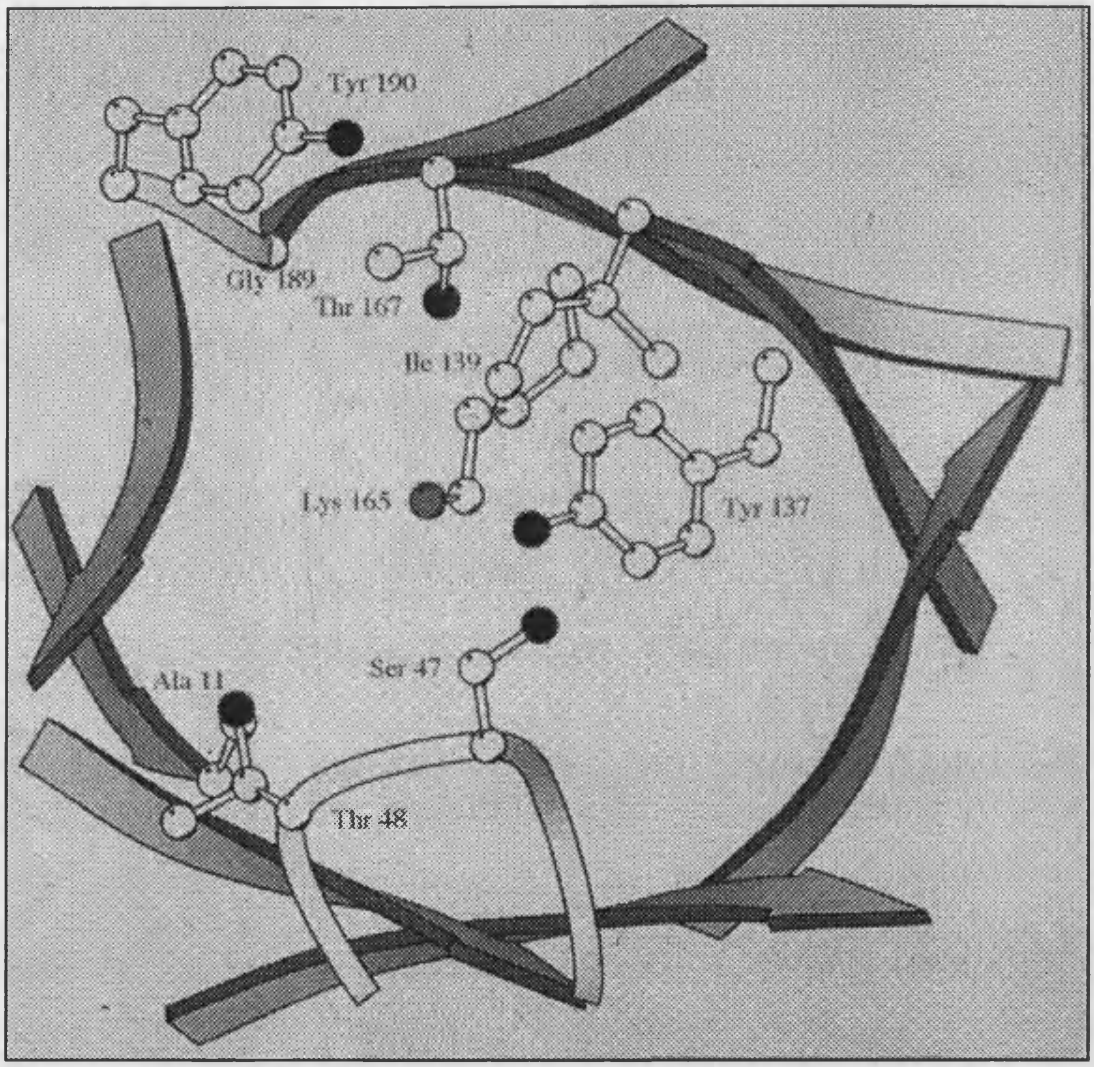


Figure 5.12 Putative active site of *E. coli* NAL. Nine of the residues lining the active site pocket are shown in ball-and-stick representation. Carbon atoms are white, oxygens black and nitrogens grey. Figure was reproduced from Izard et al. (1994) where it was created using MOLSCRIPT (Kraulis, 1991).

The Schiff-base forming lysine residue aside, it is apparent that there are other residues that play a key role in catalysis, due to their predominant conservation across the superfamily. Using the *E. coli* NAL numbering of amino acids, Thr48 is possibly involved in stabilisation of the enzyme-substrate complex, along with Ser47 (or Thr in the case of some members of the group), as they are known to form H-bonds between the NAL enzyme and inhibitor (Lawrence et al., 1997). This is compounded by Mirwaldt et al. (1995) suggesting that the high degree of conservation of residues 43 – 47 in *E. coli* DHPDS (46 – 50 in *E. coli* NAL) is due to their location in a loop region opposite the active site lysine, indicating a role in substrate binding and enzymatic activity. However, Blickling et al. (1997), found the equivalent Thr residues at positions 44 and 45 (in DHDPS) to H-bond with other amino acids within DHDPS, including Tyr133 (137 in NAL). This Tyr137 residue, conserved across all members of the superfamily, may play a role in stabilisation of the enzyme-substrate complex in that it may form a hydrogen bond with the carboxyl group of the bound pyruvate; certainly they extend towards one another in both NAL and DHPDS (Blickling et al., 1997; Lawrence et al., 1997). There is a similarly placed tyrosine (Tyr363) in human muscle fructose 1,6-bisphosphate aldolase enzyme, which also possesses an eight stranded β-barrel structure surrounded by α-helices (Gamblin et al., 1991). This tyrosine had been supposed to facilitate binding and release of the substrates dihydroxyacetone phosphate and glyceraldehyde-3-phosphate by accepting a proton from the DHAP imine C_1 and transferring it to the C_1 oxygen of G-3-P to form the C_6 imine derivative (Figure 5.13; Littlechild and Watson (1993)). This has since been shown not to be the case for this enzyme (Dalby et al., 1999), although the theory may still be relevant for KDG-aldolase. Lawrence et al. (1997) also suggests that the phenolic oxygen of Tyr137, which lies parallel to the plane of the side chain of Lys165, may contribute stability to Lys165. An intra-chain salt bridge, from Glu58 to Arg271, has also been identified in NAL, and as these residues are conserved it is supposed that this salt bridge is present in all the structures. Lawrence suggests that this salt bridge may stabilise the

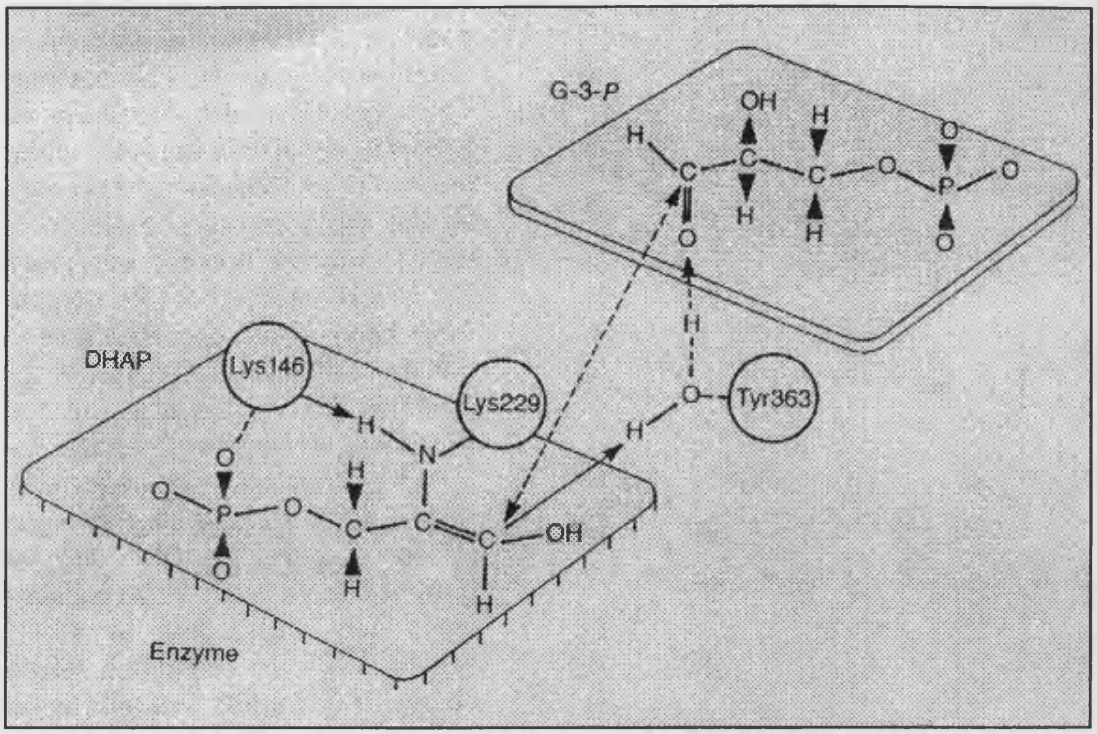


Figure 5.13 Diagrammatic representation of the aldol condensation reaction between dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (G-3-P) catalysed by fructose 1,6-bisphosphate aldolase. Lys229 forms the Schiff base with the DHAP enamine intermediate. The phenolic hydroxyl of Tyr363 accepts a proton from the DHAP imine C_1, and transfers it to the C_1 oxygen of G-3-P to form the C_6 imine derivative, thereby creating the carbon-carbon double bond as indicated by the double arrowed broken line. Figure reproduced from Littlechild and Watson (1993). This mechanism has since been shown not to be the case for this enzyme (Dalby et al., 1999), but the theory may still be appropriate to KDG-aldolase.

α-helical domain upon the surface of the barrel.

As with any predictions made based on the conservation of amino acids from sequence alignments of enzymes with known structure, these predictions cannot be supposed to be certain unless there is experimental evidence to substantiate them. The mechanism of the enzyme to proceed via a Schiff-base intermediate is experimentally determined in chapter 7. Further structural analysis is investigated by way of crystallographic studies of the enzyme (chapter 8). For a more detailed understanding of how the structure relates to the function of KDG-aldolase, further work, such as solving the three-dimensional structure of the enzyme both in the presence and absence of pyruvate is necessary, as would be site directed mutagenesis of key residues, e.g. KDG-aldolase Tyr130 → Ala. This was regrettably beyond the scope of this thesis.

6 EXPRESSION AND PURIFICATION OF RECOMBINANT ENZYME

6.1 INTRODUCTION

Having cloned and sequenced the KDG-aldolase gene, the next logical step was to express the protein recombinantly in *E. coli* in order to obtain large amounts of purified protein for characterisation studies.

As the level of expression can be affected by the choice of expression vector used, two expression systems were investigated for KDG-aldolase production. The pET System from Novagen is considered a powerful system with high levels of expression and tight control over basal expression. The target gene cloned into the pET plasmid is under the control of the T7 RNA promoter. The T7 RNA polymerase gene (λDE3 lysogen) is under the control of the *lacUV5* promoter, which has a low basal level of transcription, and is further induced by the addition of IPTG, which in turn enables the transcription of the target gene (Studier and Moffat, 1986; Rosenberg et al., 1987; Studier et al., 1990). pET-3a was selected as one potential system for high level expression.

As an alternative expression system, the pRec7 vector was also examined. pRec7 is an NdeI containing version of pARC306A (Kurz et al., 1997). This vector has a synthetic *recA* promoter, a synthetic translational enhancer region based on the sequence of the T7 promoter 10 leader, a restriction site cluster, and transcription terminators at both ends of the expression cassette. Addition of nalidixic acid to the cell culture media causes DNA damage, which activates a specific protease activity of the host cell RecA protein that cleaves and subsequently inactivates the LexA repressor of the *recA* promoter in the vector (Little and Mount, 1982). The end result of this is induction of expression of the recombinant protein.

Both of the vectors selected as potential expression systems have conveniently located NdeI and BamHI restriction sites for cloning target genes.

Having investigated the choice of expression vector, it was possible to purify KDG-aldolase for further experimentation in large amounts.

6.2 MATERIALS AND METHODS

6.2.1 Materials and reagents

Q Sepharose® Fast Flow matrix and the XK 26 column was supplied by Pharmacia Biotech, Uppsala, Sweeden.

The pRec7 plasmid was a gift from L. C. Kurz (Washington University School of Medicine, St Louis, MO, USA). The pET-3a expression vector and BL21(DE3) cells were from Novagen Inc. (Madison, WI, USA).

IPTG was supplied by Calbiochem® (Calbiochem-Novabiochem Corporation, La Jolla, CA, USA). Triton X-100 was from BDH (Poole, Dorset, UK). Lysozyme, nalidixic acid and carbenicillin were from Sigma (St. Louis, MO, USA).

6.2.2 Incorporation of cloning restriction sites into KDG-aldolase gene

PCR amplification of the KDG-aldolase gene was as described in 2.4.1 with the exceptions that 200 ng of template DNA and either 0.1 or 1 μM of each primer were used in the reaction. Primers were designed based on the codon usage Table (Table 5.1).

5' primer 5' GTTTGCATATGCCAGAA 3'

3' primer 5' TCAACCAGGATCCTATACTA 3'

6.2.3 Transformation of BL21(DE3) cells with pET-3a vector

To 1 μl and 5 μl of ligation mix in separate chilled 1.5 ml tubes 20 μl of thawed BL21(DE3) cells were added and incubated on ice for 30 min. The cells were then heat shocked at 42°C for 40 s and cooled on ice. 80 μl of LB media was added to each tube and this mixture was incubated at 37°C, with shaking, for 1 h. 50 μl of cells from each tube were spread on LB plates supplemented with 100 μg ampicillin/ml and incubated overnight at 37°C.

6.2.4 Expression of recombinant protein

LB medium supplemented with ampicillin or carbenicillin (100 μg/ml) was inoculated with an appropriate volume of LB starter culture of either BL21(DE3) cells containing the pET-3a vector with the KDG-aldolase gene, or JM109 cells containing the pRec7 vector with the KDG-aldolase gene. Cells were grown to OD_{600nm} 0.6 at 37°C, with shaking. Recombinant protein expression was induced by the addition of 0.4 mM IPTG in the case of pET-3a, or 50 μg nalidixic acid/ml in the case of pRec7. Induced cells were incubated at 25°C, 30°C or 37°C with shaking for up to 21 h.

6.2.5 Preparation of crude extract from *E. coli*

Cell pastes of various wet weights were resuspended to approximately 0.2 g/ml into 20 mM Tris-HCl, pH 8.5, 1 mM PMSF, 1 mM EDTA. Cells were lysed by incubation at 37°C for 1 h with addition of 0.1 % (v/v) Triton X-100, from a 1 % (v/v) stock made in distilled water, and 0.01 % (w/v) lysozyme, from a 1 % (w/v) stock made fresh in 20 mM Tris-HCl, pH 8.5.

Following cells lysis, DNA was sheared by four 30 s bursts of sonication on ice with a 3 mm probe, 9 – 12 μm peak to peak, using a MSE 150 watt Ultrasonic Disintegrator Mk 2 (MSE, Crawley, UK).

Soluble cell extract was obtained by centrifugation at 12,000 g for 20 min at 4°C. The pellet was discarded.

6.2.6 Heat precipitation

Soluble cell extract was heated in 1 cm diameter glass test tubes in a water bath at 78°C for 30 min. Precipitated proteins were removed by centrifugation at 12,000 g for 20 min. The pellet was discarded.

6.2.7 Q sepharose fast flow anion exchange chromatography

An XK 26 column containing 30 ml of Q Sepharose® Fast Flow anion exchange matrix was equilibrated with 20 mM Tris-HCl, pH 8.5. The supernatant, following a heat step, was applied and the column washed of unbound protein in the equilibration buffer. Bound protein was eluted at 1 ml/min with a sodium chloride gradient (0 – 1 M) in 20 mM Tris-HCl, pH 8.5, over 250 ml. 2 ml fractions were collected.

Procedures used in this chapter that are not mentioned in this section will be found in chapter 2.

6.3 RESULTS

6.3.1 Preparation of expression system

Oligonucleotide PCR primers were constructed such that they would incorporate unique restriction sites into the 5' and 3' ends of the KDG-aldolase gene (Figure 6.1). A PCR product of the correct size, 910 bp (882 bp gene, plus primer extensions of 8 bp (5') and 20 bp (3')), was produced (Figure 6.2) and then digested with NdeI and BamHI to produce 'sticky ends' or overhangs for ligation into the similarly digested expression vectors, pRec7 and pET-3a (Figure 6.3 and Table 6.1).

DNA sample	Size before digestion (bp)	Fragment size(s) after digestion (bp)
KDG-aldolase	910	892 (8, 10)
PET-3a	4,604	4,564 (40)
PRec7	2,519	2,490 (29)

Table 6.1 DNA fragment sizes before and after digestion. Digested fragment sizes indicated in brackets would not be expected to be visible on 1 % agarose gels.

Successful transformation of either expression vector into *E. coli* JM109 cells was selected for by growth on LB plates, containing ampicillin. Growth confers presence of ampicillin resistance, which would only occur due to the presence of an ampicillin resistant plasmid. Successful ligation of the KDG-aldolase gene into each plasmid was confirmed by restriction digest of DNA from single colonies. Double digests with BamHI and NdeI gives two bands on agarose electrophoresis: one for the plasmid, and one for the gene (Figure 6.4).

pET-3a containing the KDG-aldolase gene was subsequently sub-cloned into the expression host BL21(DE3), and checked by digestion as previously (Figure 6.5).

(i) KDG-aldolase gene within lambda library

```
5′  GTTTGAGGATGCCAGAA·····TAGTATACTATCATGGTTGA  3′
3′  CAAACTCCTACGGTCTT·····ATCATATGATAGTACCAACT  5′
```

(ii) Primers with incorporated restriction sites

```
5′  GTTTGCATATGCCAGAA·····                      3′
3′  ·····ATCATATCCTAGGACCAACT                   5′
```

(iii) PCR product: KDG-aldolase gene within unique restriction sites

```
5′  GTTTGCATATGCCAGAA·····TAGTATACTATCATGGTTGA  3′
3′  CAAACGTATACGGTCTT·····ATCATATCCTAGGACCAACT  5′
```

(iv) Restriction enzyme recognition sites

```
NdeI:  5′  CA | TA  TG  3′        BamHI:  5′  G | GATC  C  3′
       3′  GT  AT | AC  5′                 3′  C  CTAG | G  5′
```

(v) KDG-aldolase gene after restriction digestion

```
5′      TATGCCAGAA·····TAGTATAG          3′
3′        ACGGTCTT·····ATCATATCCTAG      5′
```

Figure 6.1 Introduction of unique restriction sites either side of KDG-aldolase gene to enable sub-cloning into an expression vector. Underlined bases indicate the location of the start and stop codons. Bases indicated in bold type are those that are altered by PCR. The centre region of the gene is illustrated by · · · · .

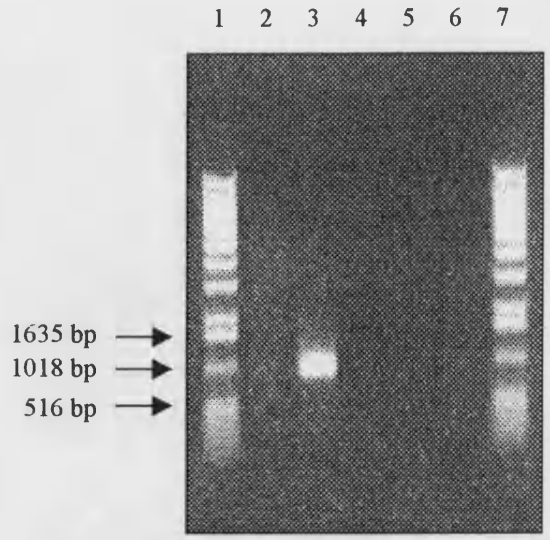


Figure 6.2 Incorporation of unique restriction sites either end of KDG-aldolase gene. Agarose gel electrophoresis enables positive identification of the PCR product (lane 3). Lane 1, DNA size markers; lane 2, PCR reaction with 0.1 μM primers; lane 3, PCR reaction with 1 μM primers; lane 4, no template DNA; lane 5, no 5' primer; lane 6, no 3' primer.

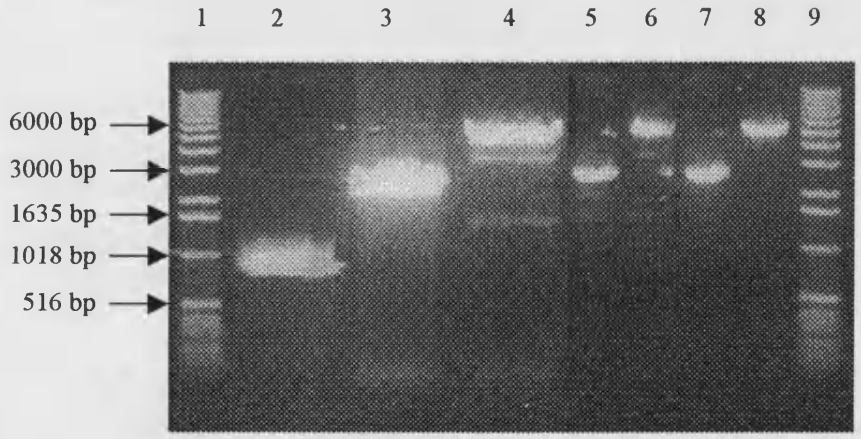


Figure 6.3 NdeI/BamHI digests of KDG-aldolase and cloning vectors. A double digest of KDG-aldolase (lane 2) gives a product minus primer ends. Double digest of pRec7 (lane 3), and pET3a (lane 4) yields linearised plasmid, minus the minor fragment of the region between the two restriction sites. A single digest of pRec7 with BamHI (lane 5), pET3a with BamHI (lane 6), pRec7 with NdeI (lane 7), and pET3a with NdeI (lane 8) gives linearised, complete plasmid.

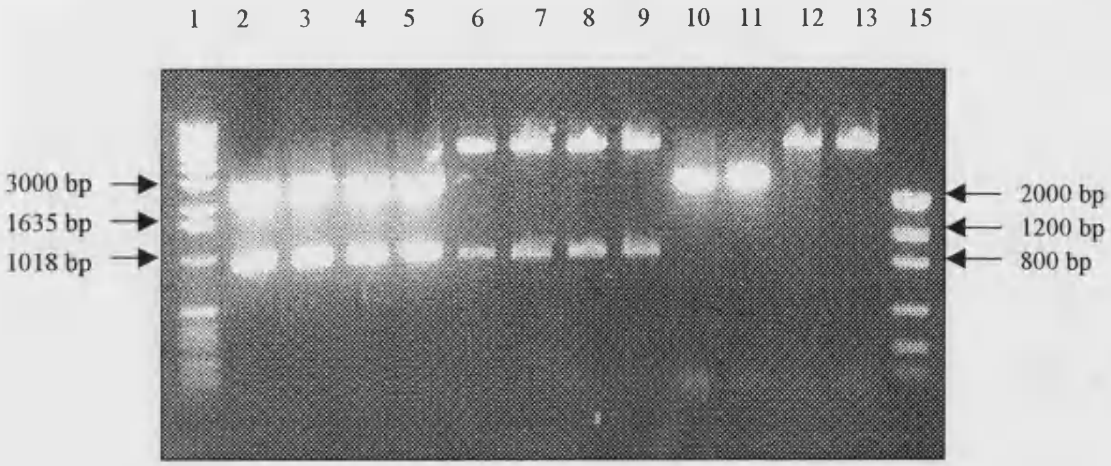


Figure 6.4 Confirmation of clone viability (1). The agarose gel electrophoresis shows the presence of the appropriate plasmid and KDG-aldolase gene, or plasmid only, in each JM109 colony. Lane 1; DNA size markers; lanes 2-5, colonies with pRec7/KDG-aldolase; lanes 6-9, colonies with pET3a/KDG-aldolase; lanes 10 and 11, colonies with pRec7 only; lanes 12 and 13, colonies with pET3a only; lane 14, DNA mass ladder. All colonies had DNA extracted and digested with NdeI and BamHI.

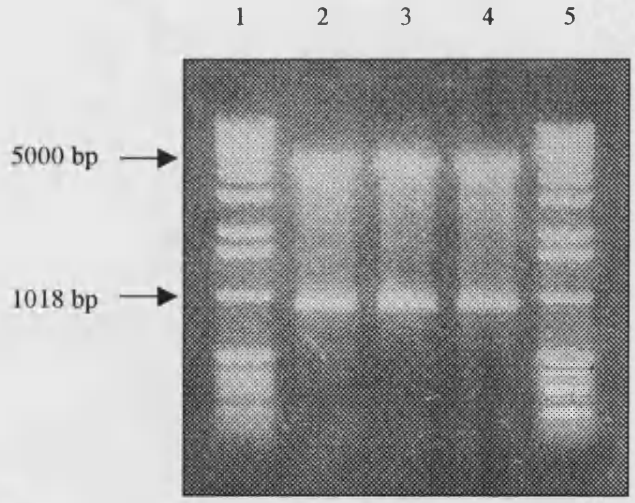


Figure 6.5 Confirmation of clone viability (2). The agarose gel electrophoresis shows presence of pET3a and the KDG-aldolase gene in each BL21(DE3) colony. Lane 1; DNA size markers; lanes 2-4, colonies with pET3a/KDG-aldolase; lane 5, DNA size marker. All colonies had DNA extracted and digested with NdeI and BamHI.

6.3.2 KDG-aldolase expression

The variability of expression with time, temperature and different expression vector was assessed. In each case, cultures of both JM109 and BL21(DE3) cells transformed with the corresponding plasmid were grown at 37°C until they had reached an OD_{600nm} of 0.6 units. Following this, half were induced with IPTG (pET-3a/KDG-aldolase) or nalidixic acid (pRec7/KDG-aldolase) and incubated at 25°C, 30°C or 37°C. Samples were taken from each culture at 0, 2, 3, 4, 7.5 and 21 h after OD_{600nm} 0.6, and analysed by SDS-PAGE (Figure 6.6a). It was apparent in both the cases of pET-3a and pRec7 that addition of IPTG and nalidixic acid respectively made little or no difference to the expression of KDG-aldolase, which was constitutively expressed. In both cases a longer incubation time after induction gave higher expression of KDG-aldolase.

The 21 h cell samples were lysed and soluble protein extract obtained for SDS-PAGE analysis (Figure 6.6) and KDG-aldolase activity assays (Figure 6.7). The level of KDG-aldolase activity when expressed using the pET-3a vector and BL21(DE3) cells is highest when induction occurs at 37°C, although expression still occurs at the lower temperatures of 25°C and 30°C. In the case of the pRec7 vector there is minimal expression at the lower temperatures; the higher temperature of 37°C is necessary for successful induction of expression. From the total activity data, it is apparent that in all cases the majority of the expressed KDG-aldolase protein is recoverable in the soluble fraction, with the largest yield occurring with expression in the pRec7 / JM109 system, with induction at 37°C over a 21 h time span.

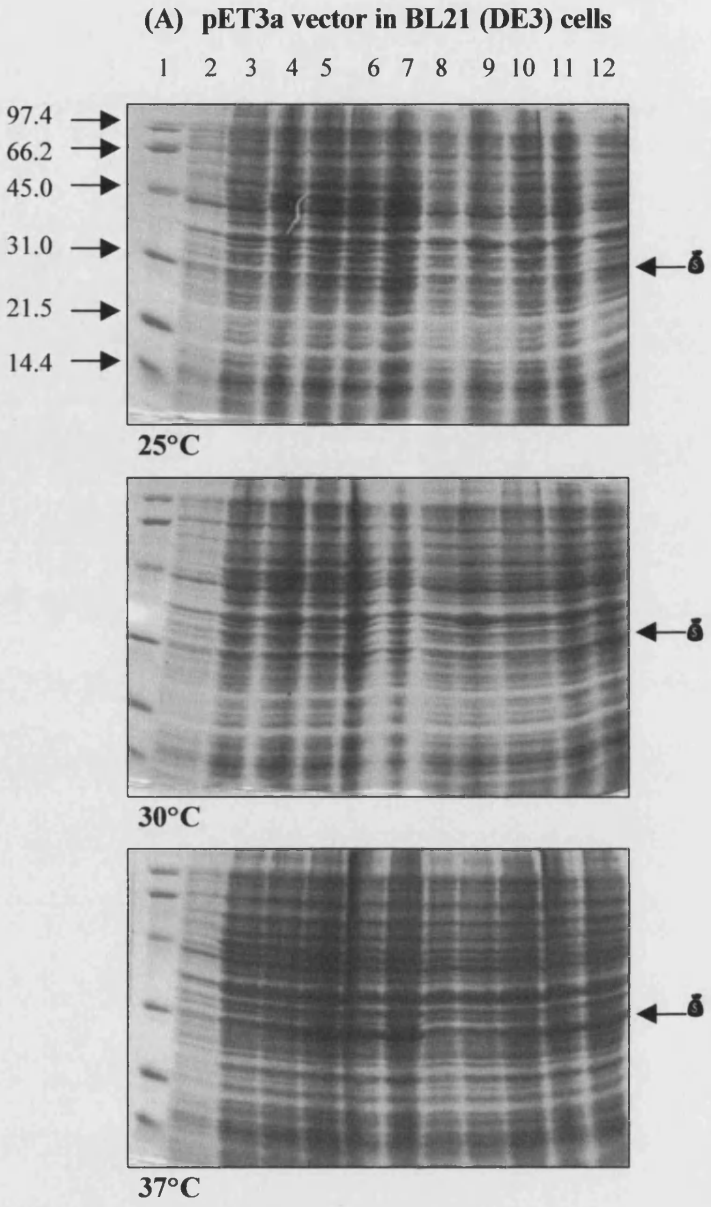
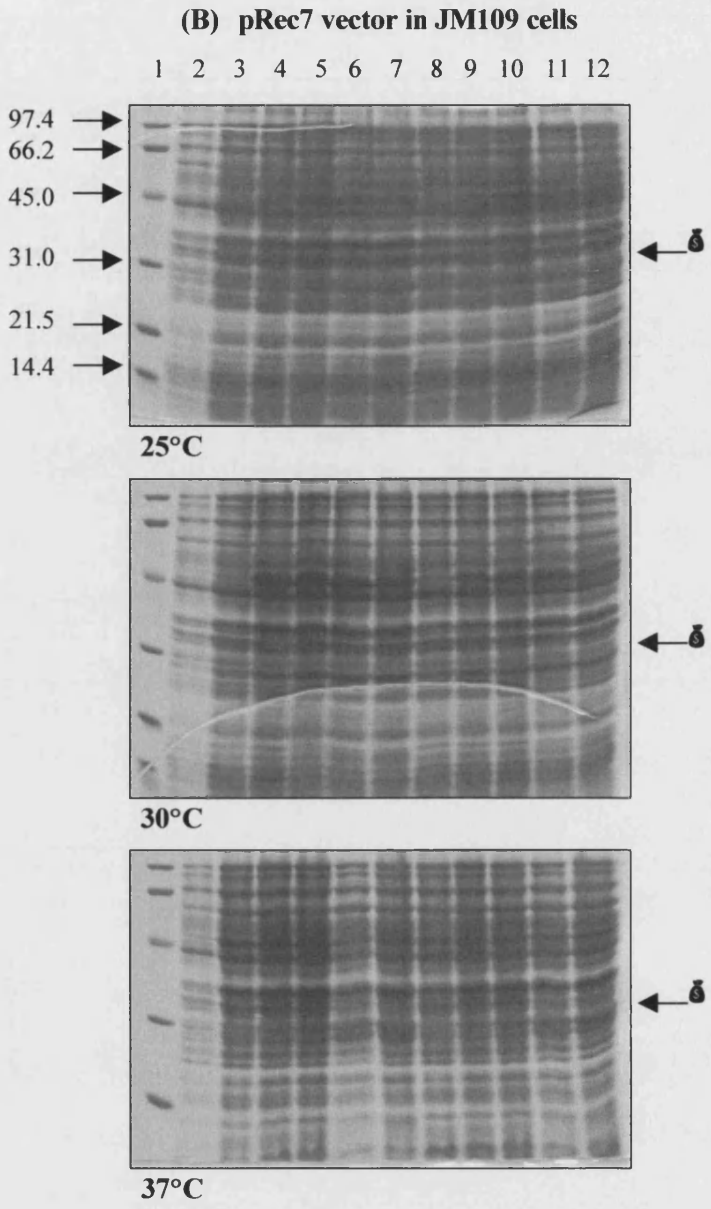


Figure 6.6a SDS-PAGE analysis of KDG-aldolase expression at different temperatures. KDG-aldolase expression using the pET3a vector in BL21(DE3) cells (A) and pRec7 vector in JM109 cells (B) was analysed following incubation at different temperaures for various times either with or without induction. For each gel, lane 1, molecular weight markers (sizes given in kDa.); lane 2, uninduced cells; lane 3, uninduced cells after 2 h; lane 4, uninduced



cells after 3 h; lane 5, uninduced cells after 4 h; lane 6, uninduced cells after 7.5 h; lane 7, uninduced cells after 21 h; lane 8, induced cells after 2 h; lane 9, induced cells after 3h; lane 10, induced cells after 4 h; lane 11, induced cells after 7.5 h; lane 12, induced cells after 21 h; . KDG-aldolase is indicated by δ .

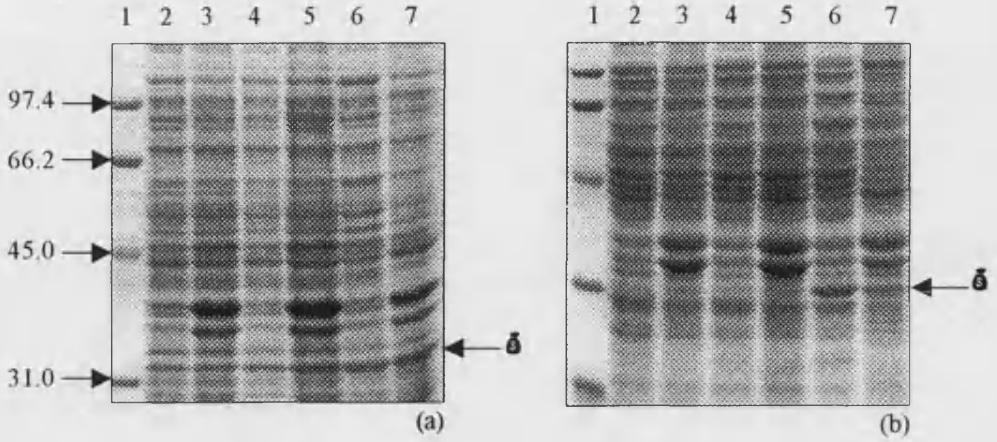


Figure 6.6 SDS-PAGE analysis of KDG-aldolase expression. Relative amounts of KDG-aldolase expression for pET3a vector in BL21(DE3) cells (a); and pRec7 vector in JM109 cells (b) was analysed. For both (a) and (b), lane 1, molecular weight markers (sizes given in kDa.); lane 2, 25°C soluble cell extract; lane 3, 25°C insoluble extract; lane 4, 30°C soluble cell extract; lane 5, 30°C insoluble extract; lane 6, 37°C soluble cell extract; lane 7, 37°C insoluble extract. KDG-aldolase is indicated by δ.

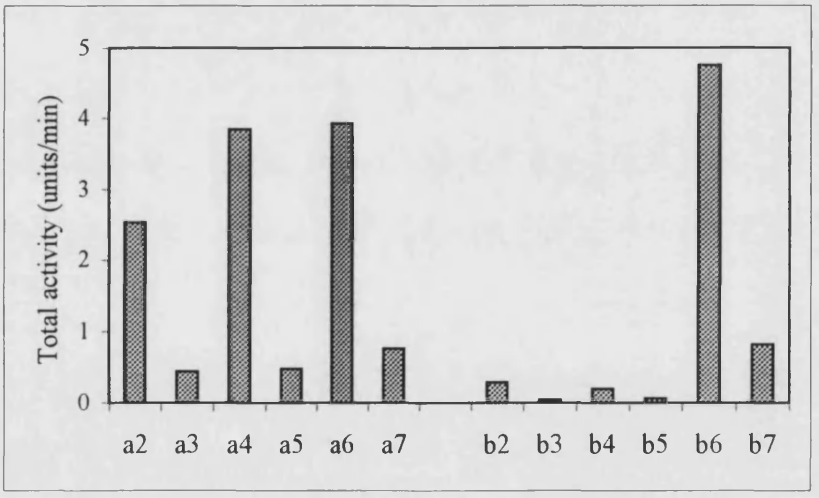


Figure 6.7 Activity analysis of KDG-aldolase expression. Relative amounts of KDG-aldolase expression for pET3a vector in BL21(DE3) cells (a); and pRec7 vector in JM109 cells (b) was analysed in the TBA assay. For both (a) and (b), bar 2, 25°C soluble cell extract; bar 3, 25°C insoluble extract; bar 4, 30°C soluble cell extract; bar 5, 30°C insoluble extract; bar 6, 37°C soluble cell extract; bar 7, 37°C insoluble extract. The bar chart order corresponds with that of the gel lanes above (Figure 6.6).

6.3.3 Recombinant KDG-aldolase purification

Recombinant KDG-aldolase was expressed in *E. coli* JM109 cells, using the expression vector pRec7, with an induction time of 21 h at 37°. Purification was achieved using a similar protocol to that found to be profitable with wild type KDG-aldolase from *S. solfataricus*. A summary of the purification results is seen in Table 6.2, with corresponding SDS-PAGE analysis displayed in Figure 6.8.

6.3.3.1 Preparation of soluble cell extract

Complete cell lysis and recovery of KDG-aldolase protein was not easily achieved by sonication and centrifugation (results not shown). For this reason an additional incubation of cells with lysosyme and Triton X-100 prior to sonication was found to be profitable. Soluble cell extract of specific activity 0.60 units/mg was obtained, which represents a 23 fold over-expression compared with the native protein in *S. solfataricus*.

6.3.3.2 Heat precipitation step

A clear benefit of expressing a thermostable protein in a mesophilic host is the ability of incubation at a high temperature to denature host cell proteins without affecting the recombinant target protein. In this instance, an incubation of 78°C for 30 min removed 80 % of contaminating proteins with a concomitant loss of only 9 % of the KDG-aldolase.

6.3.3.3 Q Sepharose® Fast Flow anion exchange chromatography

Having found anion exchange chromatography useful for the purification of native KDG-aldolase from *S. solfataricus*, the soluble fraction following the heat precipitation step was loaded onto a Q Sepharose® Fast Flow anion exchange chromatography column. Once again, the eluted fraction containing KDG-aldolase of the highest purity was retained separately, and the shoulder fractions containing less pure KDG-aldolase were pooled. The fraction with the highest specific activity of 12.45 units/mg was shown to be homogenous by SDS-PAGE and constitutes 5 % of the total cellular protein.

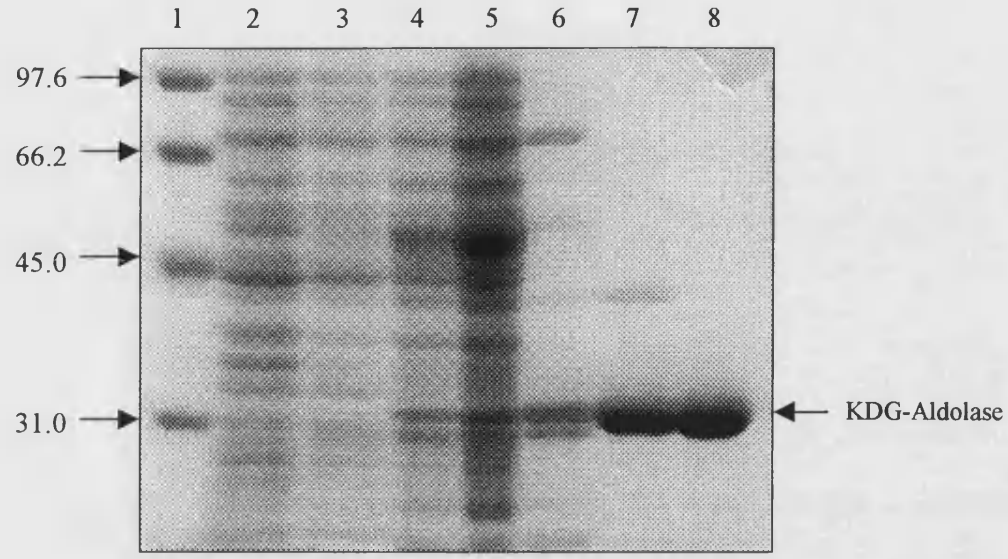


Figure 6.8 SDS-PAGE of purification of aldolase. Samples are as follows; lane 1, SDS-PAGE molecular weight standards; lane 2, non-transformed JM109 cells; lane 3, uninduced cells; lane 4, induced cells; lane 5, soluble cell extract; lane 6, supernatant after heat step and centrifugation; lane 7, Q Sepharose anion exchange pool; lane 8, Q Sepharose® Fast Flow anion exchange 'best' fraction. Molecular weights indicated by arrows are in kDa.

Purification step	Total protein (mg)	Total activity (units)	Specific activity (units/mg)	Yield (%)	Purification factor (× fold)
Soluble cell extract (5)	624	374	0.60	100	1
78°C Heat step (6)	115	338	2.94	91	5
Q Sepharose® Fast Flow anion exchange pool (7)	28	214	7.69	57	14
Q Sepharose® Fast Flow anion exchange 'best' fraction (8)	9.4	117	12.45	29	21

Table 6.2 Purification of KDG-aldolase from *S. solfataricus*. Samples after each step were analysed for activity in the TBA assay and by SDS-PAGE. Numbers in brackets after the purification step indicate the SDS-PAGE lane number in Figure 6.8. Yields and fold purification are relative to the total activity present in the soluble cell extract.

6.4 DISCUSSION

6.4.1 Expression system analyses

Having successfully cloned the KDG-aldolase gene into each of the two expression vectors, the two systems were analysed such that the more efficient and practical system could be employed for further recombinant protein production. It was apparent from initial time-course studies that both pET-3a and pRec7 vectors are 'leaky' with respect to induction control, both allowing expression of the KDG-aldolase protein in the absence of inducing agent.

Considering first the case of the pET-3a vector transformed into BL21(DE3) cells, it is known that some degree of transcription in the uninduced state is possible (Rosenberg et al., 1987). For expression of the KDG-aldolase protein, which is not expected to be in any way lethal to the host cell, this is acceptable. If expressing a protein whose affect on the cell is not innocuous, such as a protease, further control of expression would be required. For this purpose it is possible to use a host cell carrying a plasmid encoding T7 lysozyme, a natural inhibitor of T7 RNA polymerase, which introduces high stringency control, preventing transcription of target genes in the uninduced state (Studier, 1991).

The situation in the case of the pRec7 vector is more complex. The host cells used, JM109, contain a mutation in the *recA* gene rendering them *recA*$^-$. As expression of the target gene is under the control of a *recA* promoter, which is repressed by the LexA protein, the protease activity of RecA is required for inactivation of this repressor, before transcription can begin (Little and Mount, 1982; Shirakawa et al., 1984). If the host cell is unable to produce active RecA, or cannot produce RecA at all, then it would be expected that the *recA* promoter would remain repressed, and therefore no transcription to occur. However, this assumes that all of the promoters are repressed all of the time. If the amount of LexA is inadequate, or the binding constant, K_d (Equation 6.1), is sufficiently high such that there is free promoter available, then transcription would be expected to occur at a reduced rate. If transcription

does occur by this mode, then the addition of nalidixic acid would not be expected to have any effect, which was the case observed.

$$K_d = \frac{[free\ promoter][free\ LexA]}{[promoter + LexA\ complex]} \qquad \textbf{Equation 6.1}$$

It has been found previously in this lab (Gerike et al., 1997), that a *recA* mutant strain could successfully express recombinant protein from the pRec7 vector, under the control of the *recA* promoter. It was also shown that using this vector in a *recA* positive, but otherwise identical, strain would give a significantly higher level of expression.

For the purpose of producing sufficient amounts of KDG-aldolase protein for further characterisation studies, the pRec7 system, using JM109 cells, was deemed to be adequate. However, for future expression of this enzyme from this plasmid, especially for scale-up, it would be beneficial to transform the KDG-aldolase gene containing pRec7 plasmid into a *recA* positive host.

6.4.2 Recombinant KDG-aldolase purification

The purification of recombinant KDG-aldolase was achieved by a two-step method, comprising heat precipitation and anion exchange chromatography. A 29 % yield of homogenous protein, constituting 5 % of the total protein expressed, was obtained. An additional sample judged to be 95 % pure by SDS-PAGE, and representing a 57 % yield, was attained in parallel. For the scope of this study, KDG-aldolase of the highest purity was only required for crystallisation studies and relative molecular mass analysis. For this reason the pure protein was not pooled with the 'shoulder' fractions following anion exchange chromatography.

To optimise the purification of recombinant KDG-aldolase, it would be advisable to refine the heat precipitation step, following examination of the effect of different temperatures and times of incubation. Improvement of the purification factor, whilst maintaining a high yield, may increase the overall final purity and possibly reduce the requirement for a second

step in the purification protocol. Certainly, the necessity for a low-cost, time-efficient and easy purification may outweigh the benefits of highly pure protein. In this case, a one-step purification by heat precipitation could be of particular merit.

7 KDG-ALDOLASE CHARACTERISATION

7.1 INTRODUCTION

To facilitate the use of KDG-aldolase in potential large-scale biotransformations, the operational parameters of the enzyme must initially be established. This chapter describes these parameters for KDG-aldolase, and discusses them in comparison to other enzymes, some of which have been used industrially.

7.2 MATERIALS AND METHODS

7.2.1 Materials and reagents

The computer programme ENZPAK version 3.0, was supplied by BIOSOFT, Cambridge, UK.

DL-glyceraldehyde-3-phosphate was supplied as the monobarium diethylacetal derivative, by Sigma-Aldrich Company Ltd., Poole, Dorset, and prepared for use according to the manufacturers' instructions. All other reagents were also supplied by Sigma-Aldrich Company Ltd.

7.2.2 Determination of relative molecular mass (M_r)

The M_r value of the recombinant KDG-aldolase was determined by centrifugal analysis on a Beckman XL-A analytical ultracentrifuge at the UK National Centre for Hydrodynamics, University of Nottingham, UK.

Sedimentation equilibrium experiments were carried out at 20°C and three rotor speeds (10,000, 15,000 and 20,000 rpm) were employed. Three solute concentrations (0.1, 0.3 and 1 mg protein / ml) were used, and data were captured at 236 nm and 280 nm, using the scanning absorption optical system. Data were analysed using the programme NONLIN (Johnson et al., 1981), fitting to both single species and associating system models.

Sedimentation velocity experiments were carried out at 20°C and 40,000 rpm. Three solute concentrations (0.4, 0.6 and 0.8 mg protein / ml) were used, and data were analysed using the programme Svedberg (Amgen Inc., Amgen Centre, Thousand Oaks, CA, USA).

The diffusion coefficient of the KDG-aldolase was determined by Dynamic Light Scattering at 20°C on an DynaPro 801 Dynamic Light Scattering Instrument (Protein Solutions Ltd., High Wycombe, UK), using a sample at 2 mg protein / ml (Claes et al., 1992).

7.2.3 Enzyme kinetics

Kinetic analysis of the KDG-aldolase enzyme with each substrate were carried out using the TBA assay, as described in chapter 3, with minor modifications. A total reaction volume of 1 ml was used, with varying concentrations of each substrate. At various times between 0 and 10 min, 100 µl samples were withdrawn and added to 10 µl TCA. These were subsequently treated as in the standard assay. Analyses of the enzyme rates for each substrate concentration were performed using the computer programme ENZPAK, with K_m and V_{max} data being calculated according to the direct linear plot (Eisenthal and Cornish-Bowden, 1974).

7.2.4 Thermal inactivation of KDG-aldolase

Enzyme assay buffer, overlaid with mineral oil to minimise evaporation, was incubated at 100°C and 95°C in 0.5ml sealed tubes. Once the correct temperature was reached, one-fiftieth volume of recombinant enzyme was added to each tube. At selected times, a tube was removed and frozen by plunging into a dry ice / methanol bath. Tubes were stored at -20°C until assayed for KDG-aldolase activity using the TBA assay.

7.2.5 Temperature dependence of activity

Tubes containing pre-warmed 50 mM sodium phosphate buffer, pH 6.0, at 25°C, were incubated with recombinant KDG-aldolase, Na-pyruvate and DL-glyceraldehyde for 10 min at each of 30, 40, 50, 60, 70, 80, 90 and 100°C (final pH 5.99 – 5.79). Samples from each reaction were analysed for KDG produced in the TBA assay following withdrawal of a sample from each tube, and addition to one-tenth volume of 0.3 % TCA.

7.2.6 Activity in the presence/absence of divalent metal ions or EDTA

Recombinant enzyme was assayed for KDG-aldolase activity in the TBA assay either in the presence of 0.1 mM of a divalent metal ion ($CaCl_2$, $FeCl_2$, $MgCl_2$, $MnCl_2$, $ZnCl_2$,

$FeSO_4$, $MgSO_4$, $MnSO_4$, $ZnSO_4$) or 50 mM EDTA. A positive control reaction, devoid of additives, was also established.

7.2.7 Inactivation by reduction with sodium borohydride

A method similar to that used by Aisaka et al. (1991; and references therein) was used to investigate the possible involvement of a Schiff base mechanism in the KDG-aldolase reaction. Recombinant enzyme was incubated at room temperature in 50 mM sodium phosphate buffer, pH 6.0) with 100 mM $NaBH_4$ in the presence or absence of saturating concentrations of Na-pyruvate or DL-glyceraldehyde (10 mM and 20 mM, respectively). After 10 min the samples were dialysed twice against 2 l of 20 mM Tris-HCl, pH 8.5, and assayed for KDG-aldolase activity in the TBA assay.

7.2.8 Substrate specificity

Analyses of the ability of the KDG-aldolase enzyme to accept alternative substrates were performed using the TBA assay, pyruvate depletion assay, or the RI-HPLC method, each as described in chapter 3. In each case, the analogue was present in the reaction at 50 mM concentration, with the exception of DL-glyceraldehyde-3-phosphate which was used at a final concentration of 1.4 mM.

7.2.9 Inhibition of KDG-aldolase by β-hydroxypyruvate

The ability of β-hydroxypyruvate to inhibit KDG-aldolase activity was measured in the TBA assay. Reactions were set up in 250 µl total volumes, containing 50 mM, 10 mM or 1 mM Na-pyruvate, and 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5 mM β-hydroxypyruvate, such that all combinations were present. Each reaction also contained DL-glyceraldehyde and enzyme as standard in 50 mM sodium phosphate buffer, pH 7.0, and was assayed as previously described.

7.4 RESULTS

7.3.1 Determination of relative molecular mass (M_r)

Fitting of 9 data sets from the equilibrium centrifugation analysis at 3 different rotor speeds and covering a protein concentration of 0.1 to 1.0 mg / ml, demonstrated that the recombinant KDG-aldolase exists as a single species of M_r = 133,000 (± 14,000). It is concluded therefore, using M_r = 32,980 for the polypeptide translated from the sequenced gene, that the enzyme comprises 4 identical subunits, i.e. exists as a homotetramer. This was confirmed by sedimentation velocity analysis, which showed a single species of $s^0_{20,w}$ = 7.0 (± 0.3) S. The diffusion coefficient ($D^0_{20,w}$) of the KDG-aldolase was determined by dynamic laser-light scattering to be 5.4 (± 0.1) × 10^{-11} m^2/s, giving an M_r = 126,000 (± 6,000), via the Svedberg equation.

7.3.2 Enzyme kinetics

Using the TBA assay, both the native KDG-aldolase from *S. solfataricus* and the recombinant protein were found to display Michaelis-Menten kinetics, with similar catalytic properties being observed for each (Table 7.1).

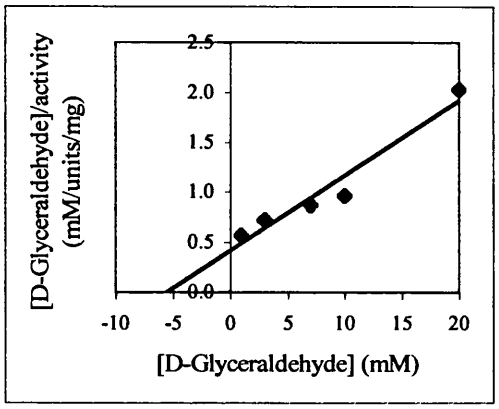
In the condensation direction, the recombinant enzyme was shown to have equal catalytic activity with both L- and D-glyceraldehyde, although their K_m values of 3.9 (± 0.3) mM and 7.1 (± 0.6) mM, respectively, were different. For assays carried out with the racemic mixture, D-glyceraldehyde will act as a competitive inhibitor of the L-glyceraldehyde, with a K_I equal to its K_m, and vice versa. In this situation, the rate equation (Equation 7.1) describing the velocity (v) observed is:

$$v = \frac{\left(V^D_{max} . S^D/K^D_m\right) + \left(V^L_{max} . S^L/K^L_m\right)}{1 + S^D/K^D_m + S^L/K^L_m} \qquad \textbf{Equation 7.1}$$

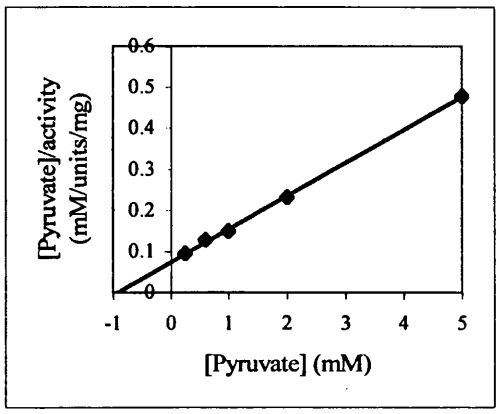
where the substrate concentrations (S) and kinetic constants (K_m and V_{max}) of the D- and L-glyceraldehyde are indicated by their respective superscripts (Wharton and Eisenthal,

	K_m (mM)	Specific activity at V_{max} (units/mg)	k_{cat} (min^{-1})
NATIVE KDG-ALDOLASE PURIFIED FROM *S. SOLFATARICUS* (SEMI-PURIFIED)			
DL-Glyceraldehyde	3.6 (± 0.2)	8.4 (± 0.2)	
Pyruvate	0.9 (± 0.2)	6.3 (± 0.2)	
RECOMBINANT KDG-ALDOLASE PURIFIED FROM *E. COLI* (PURIFIED)			
D-Glyceraldehyde	7.1 (± 0.6)	18.0 (± 1.9)	594 (± 62.7)
L-Glyceraldehyde	3.9 (± 0.3)	18.0 (± 1.0)	594 (± 33.0)
DL-Glyceraldehyde	5.2 (± 0.1)	17.1 (± 0.4)	564 (± 13.2)
Pyruvate	1.0 (± 0.1)	15.7 (± 0.3)	517 (± 9.9)
KDG	8.6 (± 1.9)	15.7 (± 2.3)	517 (± 75.8)

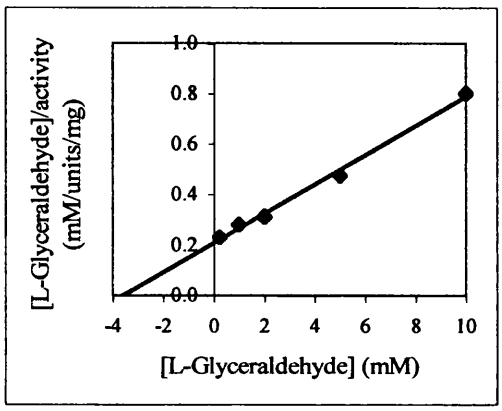
Table 7.1 Kinetic parameters for native and recombinant KDG-aldolase. Data were analysed using the Direct Linear Plot (Eisenthal and Cornish-Bowden, 1974). Values in parenthesis are standard errors. Kinetic parameters for pyruvate were obtained using 20 mM DL-glyceraldehyde. Kinetic parameters for D-, L- and DL-glyceraldehyde were obtained using 50 mM pyruvate. Each pair of results was determined from a minimum of two experiments. Typical plots are shown in Figure 7.0.



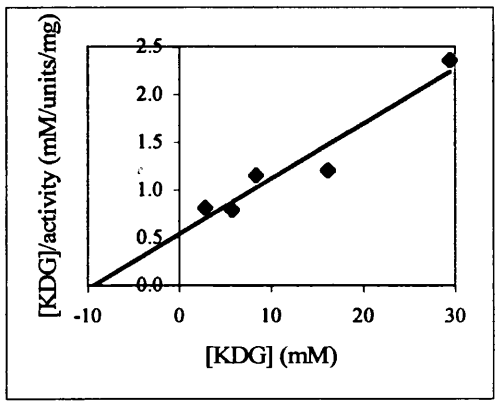
(a)



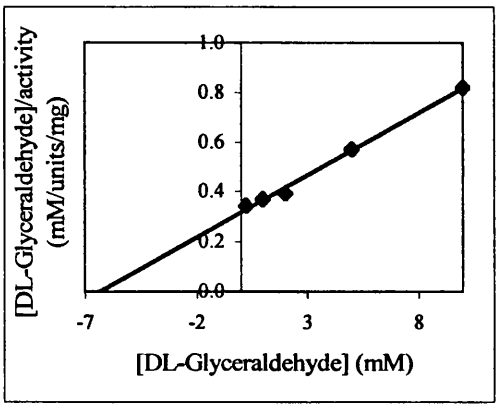
(d)



(b)



(e)



(c)

Figure 7.0 Graphs of typical results for kinetic parameters. (a) D-glyceraldehyde, (b) L-glyceraldehyde, (c) DL-glyceraldehyde, (d) pyruvate and (e) KDG.

1981). Given that the V_{max} values for the two substrates are identical, the K_m that would be determined from assays with the racemic mixture can be calculated from the rate Equation 7.2:

$$K_m = \frac{2K_m^D . K_m^L}{K_m^D + K_m^L} \qquad \textbf{Equation 7.2}$$

Inserting the values given in Table 7.1 for the individual substrates, the predicted value of K_m for the racemic mixture is calculated as 5.0 mM, compared to an experimentally determined value of 5.2 (± 0.1) mM.

7.3.3 Thermal inactivation of KDG-aldolase

In the absence of substrates, the recombinant enzyme has been demonstrated to be extremely thermostable, having a half-life of approximately 3 h at 100°C (Figure 7.1a), and approximately 6 h at 95°C (Figure 7.1b), as measured by the rate of irreversible thermal inactivation. Although a straight line has been fitted to the data in Figure 7.1b for the convenience of the half-life determination, it is noted that the thermal inactivation may occur via a more complex, possibly biphasic, pattern.

7.3.4 Temperature dependence of activity

Recombinant KDG-aldolase is active over a wide range of temperatures (Figure 7.2a), with activity approximately doubling with every 10°C rise in assay temperature between 30°C and 80°C. The optimum temperature is 93°C. An Arrhenius plot can be constructed, according to the Arrhenius expression (Equation 7.3), of ln(velocity) against $1/T$, giving a straight line of slope $-E_a/R$ (Price and Stevens, 1989). The E_a value is 1.28 J^{-1} K mol. These data are linear over this temperature range (Figure 7.2b), but a break is observed at 80°C. However, as indicated from the thermostability data described above, the lower than expected activities above 80°C are not due to inactivation of the enzyme, as only 3 % loss of activity is expected at 100°C over the 10 minute period of the assay (Figure 7.1a). It has also been shown that there is no significant decomposition of the substrates over 10 min

at these temperatures. Both enzyme and substrates are therefore shown to be extremely thermostable under normal assay conditions.

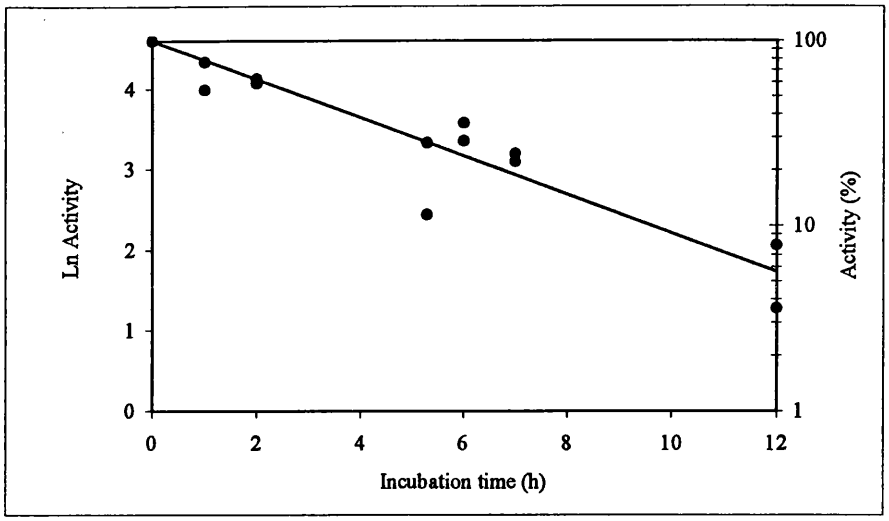


Figure 7.1a Thermal inactivation of KDG-aldolase at 100°C. Purified recombinant KDG-aldolase was incubated at 100°C for up to 12 h, as described in section 7.2.4. After rapid cooling, residual enzymatic activity was measured at each time point in the TBA assay.

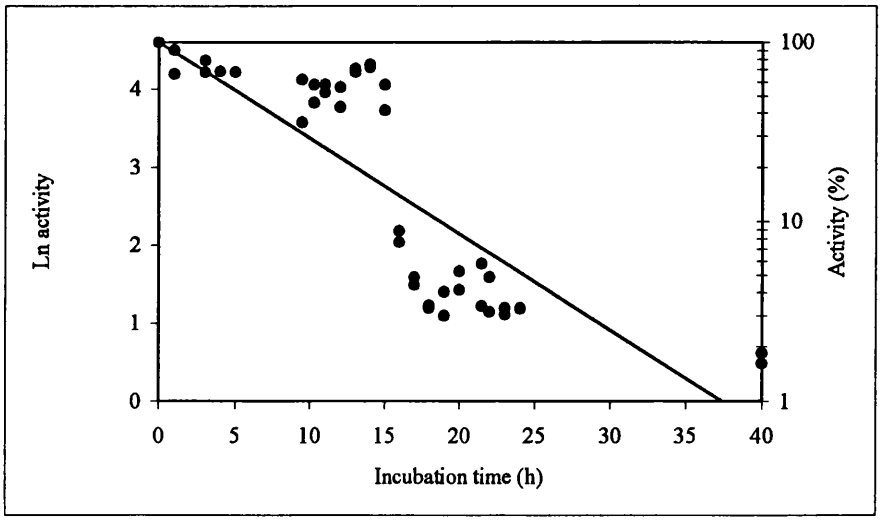


Figure 7.1b Thermal inactivation of KDG-aldolase at 95°C. Purified recombinant KDG-aldolase was incubated at 95°C for up to 40 h, as described in section 7.2.4. After rapid cooling, residual enzymatic activity was measured at each time point in the TBA assay.

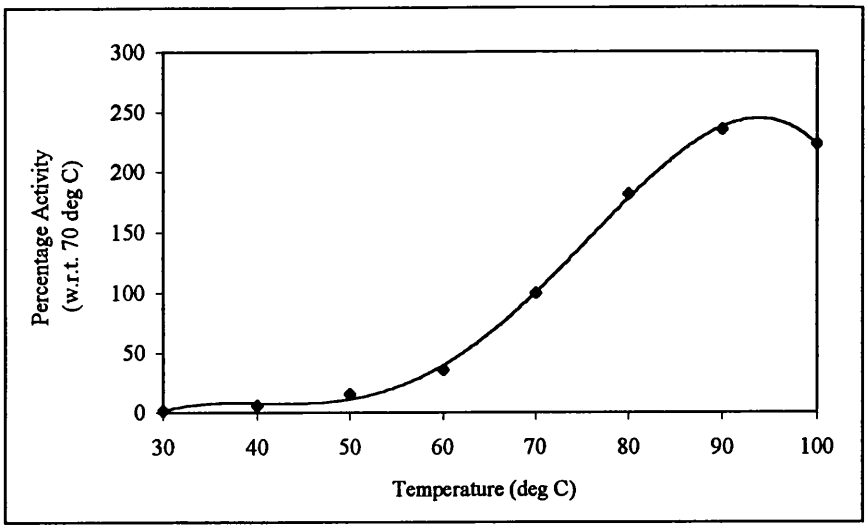


Figure 7.2a Temperature-dependence of KDG-aldolase activity. Purified recombinant KDG-aldolase was assayed as described in section 7.2.5 at temperatures from 30 - 100°C. Activities are expressed as a percentage of the activity at 70°C, the routine assay temperature.

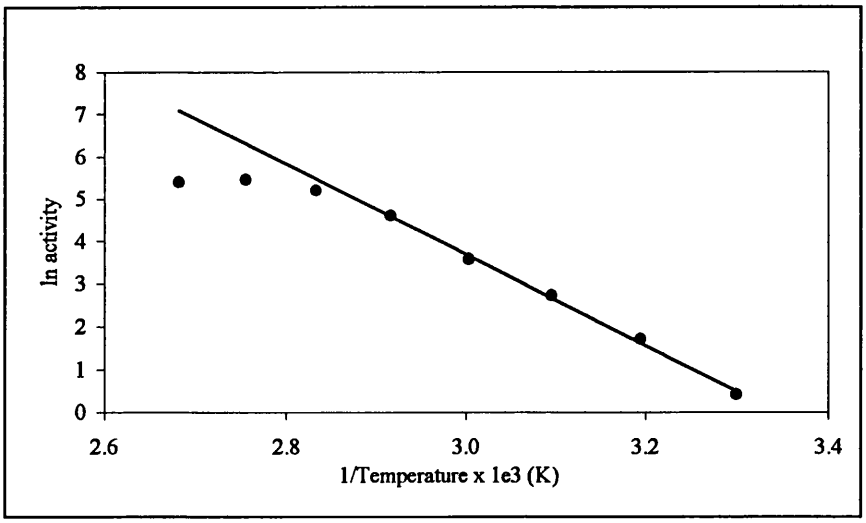


Figure 7.2b Arrhenius plot. Data from figure 7.2a is illustrated in the form of the Arrhenius plot, according to the Arrhenius expression shown in Equation 7.3.

$$k = Ae^{-E_a/RT} \qquad \textbf{Equation 7.3}$$

where A is known as the pre-exponential factor,

R is the gas constant

T is the absolute temperature, and

E_a is the activation energy for the reaction.

7.3.5 Activity in the presence/absence of divalent metal ions or EDTA

Aldolase enzymes can be classified as either Type I enzymes, which form a Schiff-base intermediate with the donor substrate via an active site lysine residue, or as Type II enzymes, which require a metal ion cofactor (Rutter, 1964). In the standard TBA assay, recombinant KDG-aldolase activity was unaffected by the addition of 0.1 mM of various divalent metal ions (Figure 7.3). Additionally, activity was not compromised by the addition of 50 mM EDTA. This is in contrast to the thermostable type II aldolase from *Methanococcus janaschii*, which was inactivated in the presence of EDTA and reactivated by addition of Zn^{2+} ions (Choi et al., 1998).

7.3.6 Inactivation by reduction with sodium borohydride

The involvement of a Schiff-base mechanism in the KDG-aldolase reaction was examined by treating the recombinant enzyme with sodium borohydride in the presence or absence of the substrates pyruvate and DL-glyceraldehyde. The results summarised in Table 7.2 indicate the participation of a Schiff-base mechanism in catalysis; 69 % inactivation of the enzyme was observed following incubation with pyruvate and $NaBH_4$ for 10 min, and complete inactivation was observed when the incubation was increased to 20 min. These results, combined with those from 7.3.5, indicate that KDG-aldolase can be classed as a Type I aldolase enzyme.

7.3.7 Substrate specificity

The ability of KDG-aldolase to utilise alternative substrates to pyruvate and glyceraldehyde in the aldol condensation reaction was examined. The structures of the compounds tested are illustrated in Table 7.3A and B along with a summary of the results observed.

Each of the three pyruvate analogues and the DL-Glyceraldehyde-3-phosphate were conclusively unable to react in the KDG-aldolase reaction in the place of the *in vivo* substrate.

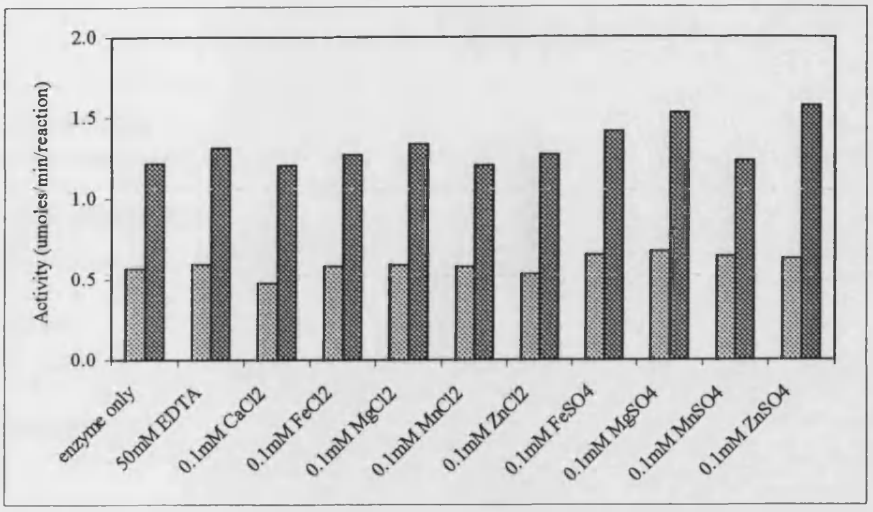


Figure 7.3 Activity of recombinant KDG-aldolase in the presence / absence of divalent ions or EDTA. Each assay was carried out in duplicate with 25 µl (▨) or 50 µl (▨) KDG-aldolase added.

INCUBATION CONDITIONS	ACTIVITY REMAINING (%)
Enzyme	100
Enzyme + $NaBH_4$	78
Enzyme + $NaBH_4$ + pyruvate	31
Enzyme + $NaBH_4$ + DL-glyceraldehyde	90

Table 7.2 Inactivation by sodium borohydride reduction. Recombinant KDG-aldolase was incubated with $NaBH_4$, for 10 min, in the presence and absence of the substrates pyruvate and DL-glyceraldehyde.

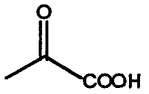
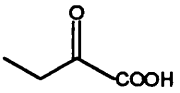
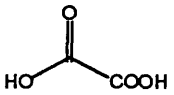
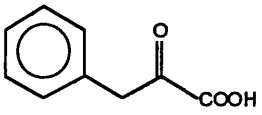
NAME OF COMPOUND	STRUCTURE OF COMPOUND	ASSAY METHOD	RESULT
A. PYRUVATE ANALOGUES			
α-Ketobutyrate		TBA	no activity
β-Hydroxypyruvate		TBA	no activity
Phenylpyruvic acid		RI-HPLC	no activity

Table 7.3A Substrate specificity of KDG-adolase with pyruvate analogues. For each potential substrate is given the structure, the assay in which it was analysed, and the result of the enzyme activity assay using the alternative substrate in place of pyruvate

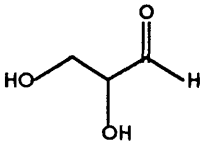
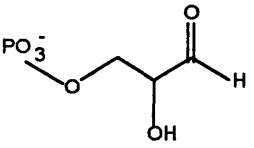
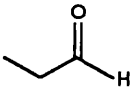
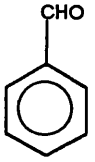
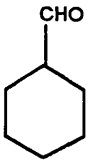
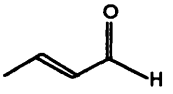
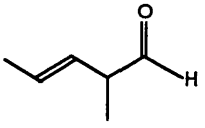
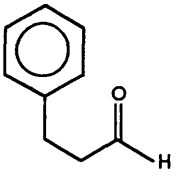
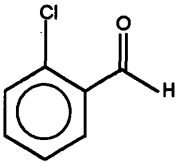
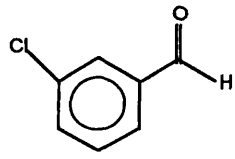
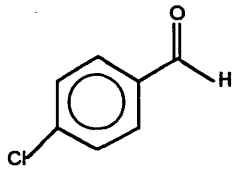
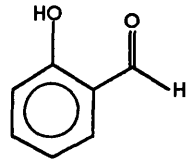
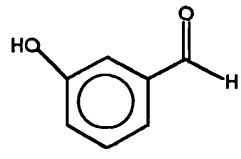
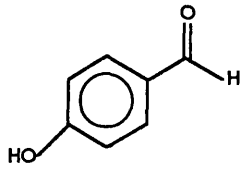
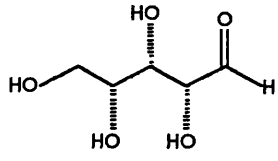
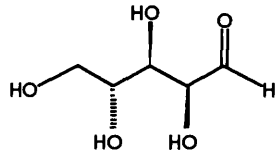
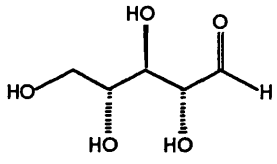
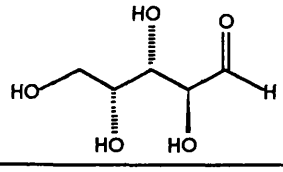
NAME OF COMPOUND	STRUCTURE OF COMPOUND	ASSAY METHOD	RESULT
B. GLYCERALDEHYDE ANALOGUES			
DL-Glyceraldehyde-3-phosphate		TBA	no activity
Propionaldehyde		PD	++
Benzaldehyde		PD	+
Cyclohexanecarboxaldehyde		PD	++
Crotonaldehyde (2-butenal)		PD	+++
Trans-2-methyl-2-butenal		PD	++
Phenylacetaldehyde		PD	+++
2-Chlorobenzaldehyde		PD	+

Table 7.3B Substrate specificity of KDG-aldolase with glyceraldehyde analogues. For each potential substrate is given the structure, the assay in which it was analysed, and the result of the enzyme activity assay using the alternative substrate in place of glyceraldehyde. An overall assessment of activity relative to that with DL-glyceraldehyde is given for the glyceraldehyde analogues. N.B. The pyruvate depletion assay is indicated by the abbreviation PD.

NAME OF COMPOUND	STRUCTURE OF COMPOUND	ASSAY METHOD	RESULT
3-Chlorobenzaldehyde		PD	++
4-Chlorobenzaldehyde		PD	+
2-Hydroxybenzaldehyde (salicyldehyde)		PD	+
3-Hydroxybenzaldehyde		PD	++
4-Hydroxybenzaldehyde		PD	+++ / ?
D-Ribose		PD	+
D-Lyxose		PD	++
D-Xylose		PD	+
D-Arabinose		PD	++

Key: + = 0 – 16 %, ++ = 17 – 30 %, +++ = > 30 %

(values are percentage activities, relative to the maximum activity with DL-glyceraldehyde)

The results for the glyceraldehyde analogues analysed in the pyruvate depletion assay were not conclusive due to their inconsistency. Analysis of the results for the glyceraldehyde analogues was by an approximation of two experiments where reaction samples were taken and analysed at intervals up to 6 h. In neither case did the activity with the analogue reach more than 50 % of that with DL-glyceraldehyde. Analysis with 4-hydroxybenzaldehyde was particularly challenging due to the variability of the high background absorbance of this compound at 320 nm. It was therefore difficult to establish whether there was significant activity with this compound.

7.4.8 Inhibition of KDG-aldolase by β-Hydroxypyruvate

As well as testing KDG-aldolase for activity with β-hydroxypyruvate as an alternative substrate to pyruvate, the activity was measured with varying concentrations of β-hydroxypyruvate in the presence of pyruvate. β-Hydroxypyruvate was found to be a competitive inhibitor, such that the V_{max} of the enzyme remained unchanged, and the apparent K_m for pyruvate increased. This result is shown in Figure 7.4a. The mean of the x co-ordinates of the intersections of the lines gives the inhibition constant, K_I, to be 0.39 (± 0.03) mM. The parallel lines in the Cornish-Bowden plot (Figure 7.4b) confirm that the inhibition is pure competitive.

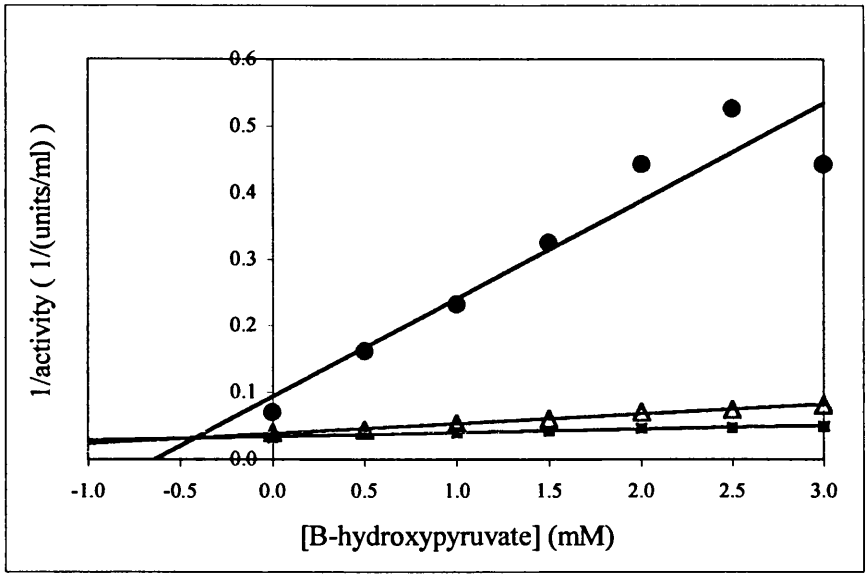


Figure 7.4a Dixon plot illustrating inhibition of KDG-aldolase activity by β-hydroxypyruvate. KDG-aldolase activity was measured for varying β-hydroxypyruvate concentration, in the presence of 1 mM (•), 10 mM (Δ) and 50 mM (■) pyruvate. A K_I of 0.39 (± 0.03) mM, is given by the intersection of the lines.

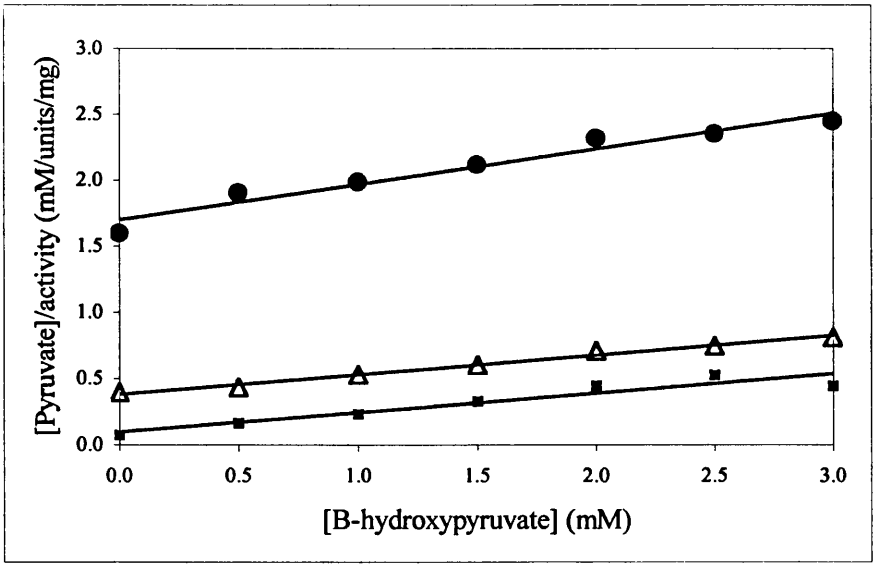


Figure 7.4b Cornish-Bowden plot illustrating competitive inhibition of KDG-aldolase activity by β-hydroxypyruvate. KDG-aldolase activity was measured for varying β-hydroxypyruvate concentration, in the presence of 1 mM (•), 10 mM (Δ) and 50 mM (■) pyruvate.

7.4 DISCUSSION

7.4.1 Determination of relative molecular mass (M_r)

KDG-aldolase has been shown in this study to exist as a homotetramer, and in this respect it is similar to the enzymes NAL and DHDPS, which are also tetrameric. Interestingly, the light scattering data indicate that the KDG-aldolase protein is roughly discoidal in morphology, a similar trait having been observed for NAL (Izard et al., 1994), and which has been suggested to account for the misinterpretation of enzymes from this superfamily being initially identified as trimeric (Uchida et al., 1984).

7.4.2 Enzyme kinetics

The comparison of DL-glyceraldehyde and pyruvate K_m values for the native and recombinant enzymes supports the claim that the two enzymes are identical and no erroneous modifications of the recombinant enzyme have taken place.

The large standard errors for the KDG K_m and V_{max} data are not unexpected, as substrate depletion rather than product formation was measured in the TBA assay. This comparatively small change to be measured introduces additional error to the experiment, as has been observed when using the pyruvate-depletion assay (section 3.3.2) where the standard error was 5.6 %.

7.4.3 Thermal inactivation and temperature dependence of KDG-aldolase activity

Experiments at elevated temperatures over varying time periods have shown KDG-aldolase to have a high optimum temperature of activity of 93°C with remarkable thermostability above this temperature. Incubation over short time periods above 90°C indicate that the enzyme undergoes a reversible dissociation or unfolding that affects the catalytic activity. If incubated for extended periods of time, then an irreversible denaturation

occurs. The observation of this effect has also been proposed for *Thermoanaerobacter* 3-phosphoglycerate kinase (Thomas and Scopes, 1998) and emphasises the point that thermostability alone does not guaranteed thermoactivity. It is interesting to note that Thomas and Scopes (1998) also made the point that as the temperature increases a concomitant increase in K_m values is also frequently observed. This leads to an underestimation of K_m values if the substrates are not present in saturating concentrations. In this instance it is unlikely to be the case however, as the linearity of the reaction was assessed for each temperature tested. Ideally, the enzyme efficiency (k_{cat}/K_m) should be established for each reaction temperature.

The advantages of enzyme thermostability in biotechnology include factors such as allowing reactions to proceed at high temperatures that may be necessary for substrate solubility. When analysing various aldehyde substrates in the KDG-aldolase reaction it was observed that many of them are not soluble at room temperature, and therefore the high temperature of the assay was profitable. This benefit is harnessed in the enzymatic hydrolysis of biopolymers, such as starch (Nigam and Singh, 1995), because of the improved solubility and the reduction in viscosity at high temperatures.

A further advantage of thermostable enzymes is their apparent resistance to proteolysis and irreversible chemical degradation, such as denaturation by organic solvents, at high temperatures (reviewed by Sellek and Chaudhuri, 1999). This characteristic may be due to the increased rigidity of thermostable enzymes, governed by any combination of various improvements made in the enzyme's packing efficiency, such as a reduction in internal cavity size and thermolabile residue content or increased number of ion pairs and hydrophobic inter-subunit interactions (Russell et al., 1997). An example is the resistance of GAPDH from *P. woesei* to deamidation and peptide bond hydrolysis correlating with the enzyme's increased conformational stability (Hensel et al., 1992).

7.4.4 Schiff-base reaction mechanism

The ability of the KDG-aldolase reaction to proceed in the absence of metal ions and to be unaffected by their addition or the presence of EDTA, but to be inactivated following sodium borohydride treatment in the presence of pyruvate, all suggests that this enzyme is a type I aldolase. Type I aldolases proceed via Schiff-base formation between the carbonyl group of pyruvate and the ε-amino group of a lysine residue at the active site of the enzyme. This lysine residue has been identified in the *E. coli* enzymes NAL and DHPDS as Lys165 and Lys161 respectively (Aisaka et al., 1991; Laber et al., 1992) and is conserved throughout the NAL superfamily (Figure 5.5).

A proposed scheme for the reaction mechanism of the KDG-aldolase is given in Figure 7.5. In the direction of aldol condensation, the carbonyl group of pyruvate is subjected to nucleophilic attack by the ε-amino group of Lys155, resulting in the formation of the imine. The phenolic oxygen of Tyr120 (Tyr133 in *E. coli* DHPDS), the oxygen of the peptide linkage of Val 196, or a water molecule, is then postulated to act as the proton acceptor/donor from the C_1 of the imine to the C_1 oxygen of the glyceraldehyde, forming the C_6 imine derivative. Release of the C_6 imine gives KDG. An analogous reaction mechanism had been proposed for human muscle fructose-bisphosphate aldolase (Littlechild and Watson, 1993), although the tyrosine residue originally suggested to act as the proton acceptor/donor has since been found not to be involved in the reaction mechanism (Dalby et al., 1999). In the case of *E. coli* DHDPS, Tyr133 (Tyr120 in KDG-aldolase) has been suggested to assist the reaction by way of a hydrogen bond to the keto oxygen of pyruvate, due to its proximity to the pyruvate C_2. The backbone oxygen of Ile 203 (Val 196 in KDG-aldolase), which is conservatively exchanged across the NAL superfamily, may assist in proton transfer; alternatively, a water molecule may be the mediator (Blickling et al., 1997).

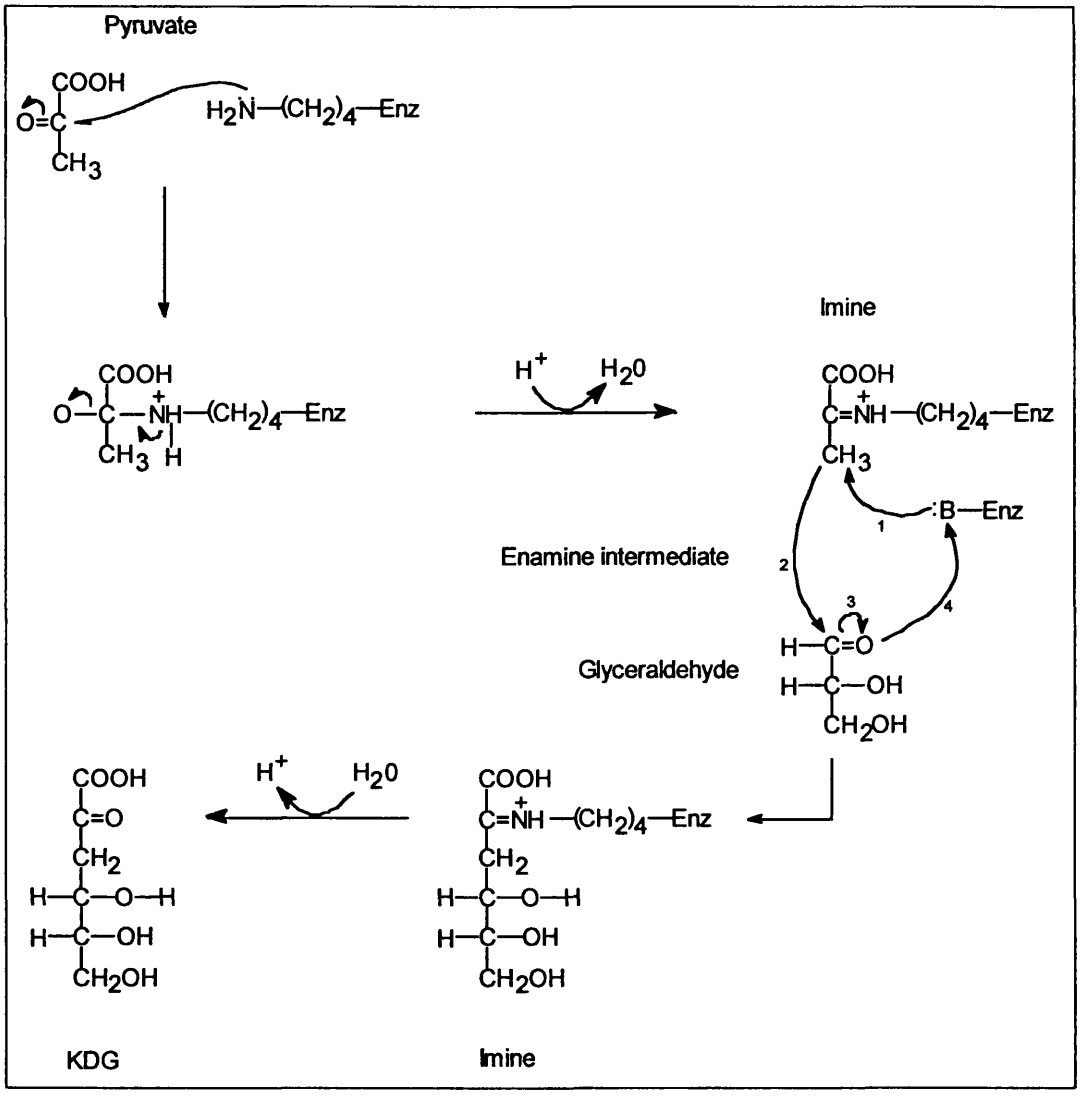


Figure 7.5 The proposed mechanism for the KDG-aldolase reaction. The carbonyl group of pyruvate is subjected to nucleophilic attack by the ε-amino group of Lys155, forming the imine derivative. A proton is then passed from C_1 of the imine, via a suitable base (possibly Tyr120, V196 or H_2O) to the C_1 oxygen of glyceraldehyde, resulting in the formation of a C_6 imine derivative, via a transient enamine intermediate. The transfer of electrons is indicated by arrows, and is labelled 1,2,3,4 to indicate the order of transfer. Finally, the C_6 imine is released from the Schiff-base intermediate to yield KDG.

7.4.5 Substrate specificity

Detailed understanding of the substrate specificity of KDG-aldolase has two benefits. Knowledge of the chemical structure of acceptable substrates may allow the piecing together of predictions that can be made concerning the active site environment of the enzyme, leading to a more thorough comprehension of the reaction mechanism itself. The second benefit is that an enzyme with a large repertoire of potential substrates has a similarly large repertoire of potential products, which would make the enzyme an attractive candidate for the biosynthesis of novel compounds.

It is unfortunate that the information available for KDG-aldolase is limited by the assay methods available. However, it has been shown that KDG-aldolase appears to be specific in the case of pyruvate, as neither α-ketobutyrate nor β-hydroxypyruvate give detectable products in the TBA assay. Similar specificity for the keto-acid is observed with *E. coli* NAL (Uchida et al., 1984; Kim et al., 1988) and *Aureobacterium barkerei* 3-deoxy-d-*manno*-2-octulosonic acid (KDO) aldolase (Sugai et al., 1993). Flexibility is however observed with these enzymes for the electrophilic acceptor. KDG-aldolase appears to accept a range of aldehyde substrates, although it is not possible to draw conclusions concerning structural parameters governed by the enzyme, as the data were so variable. Further development of the RI-HPLC assay may be the most productive method for specificity studies. HPLC methods have previously been successfully used to analyse the specificity in the case of various aldolase enzymes (Lotz et al., 1990; Pollard et al., 1998).

8 KDG-ALDOLASE CRYSTALLISATION STUDIES

8.1 INTRODUCTION

8.1.1 What is X-ray crystallography and how does it work (Rhodes, 1993).

The first three-dimensional structure of a protein to be solved at near atomic resolution was sperm whale myoglobin by Kendrew et al. (1960). Since then, many more structures have been solved, and this information has been used as a valuable aid to structure-function relationship studies.

The technique is based on the ability of an ordered array of identical molecules, e.g. crystalline protein, to diffract light. This gives a pattern that relates to the position of the individual atoms in the crystal. Because the individual atoms in a protein molecule are approximately 1.5 Å apart, the wavelength of the light must not exceed this. Therefore, electromagnetic radiation in the X-ray range is used, giving a diffraction pattern that can be interpreted such that the position of each atom can be resolved accurately. However, it is an appreciable hurdle to the crystallographer that it is not possible to produce a focused model of the protein from the diffracted X-ray beam, as X-rays cannot be focused by lenses. Instead, computer analysis of the intensities and directions of the diffracted X-rays is required, the computer mimicking the work of an objective (focussing) lens.

Crystallisation of the protein occurs under certain conditions where there is a slow, controlled precipitation from aqueous solution. It is imperative that the conditions do not denature the protein. Factors influencing the crystallisation of a protein include temperature, pH, ionic strength and protein concentration. Finding the optimum conditions for precipitation to occur is initially based on trial and error screening.

The crystals themselves are formed of ordered molecules in only one or a few orientations held together by non-covalent interactions. The entire crystal is composed of neatly stacked unit cells, each one representing the smallest repeatable unit representative of the whole crystal. The repetitive nature of the crystal increases the intensity of the diffraction pattern.

The main obstacle besetting the crystallographer is known as the phase problem. In simple terms, the solution of the phase problem is not unlike solving simultaneous equations. Each diffracted X-ray represents amplitude, frequency and phase. The amplitude is a function of the measured intensity, the frequency is that of the X-ray source, but the phase is not directly obtainable. With so many protein crystal structures presently available, it is common to solve the phase problem by a method known as molecular replacement. This uses the phases of a known protein (the phasing model) as initial estimates of phases for the protein that the structure is being determined. The two proteins must be related structurally, and therefore have significant sequence homology. An alternative method is isomorphous replacement, where heavy-atoms are soaked into a crystal giving rise to slight perturbations in the diffraction pattern. These slight changes can be defined by one of two possible solutions, such that it is common to require two heavy-atom derivatives to solve one phase problem.

8.1.2 KDG-aldolase and X-ray crystallography

It is desirable to know the three-dimensional structure of KDG-aldolase such that the reaction mechanism of the enzyme can be understood in more depth. The crystal structures of both NAL and DHDPS have been solved, and this has allowed predictions to be made about how these enzymes orientate their substrates (Izard et al., 1994; Mirwaldt et al., 1995). A more detailed understanding of the active site of KDG-aldolase would enable a more thorough comparison of these enzymes than can be given from primary structures. Combined with specificity studies, suggestions as to how the enzyme confers substrate specificity could be made, and the rational design of mutants enabled.

The existing structural information for either the *E. coli* NAL or DHDPS enzyme provides a phasing model for solving the KDG-aldolase structure by X-ray crystallography. For this study, the phases from the *E. coli* NAL enzyme were chosen.

8.2 MATERIALS AND METHODS

8.2.1 Materials

The Crystal Screen kit, PEG 6000 Grid Screen, ammonium sulphate Grid Screen kit, 24 well Linbro plates, 22 mm circle cover slides, Prosil 28 and Dow Corning vacuum grease were all supplied by Hampton Research, Laguna Niguel, CA, USA. The FPLC Superdex 200 gel filtration column and associated apparatus was supplied by Pharmacia Biotech, Uppsala, Sweden. Centricon 30 concentrators were from Amicon, Inc., Beverly, MA, USA.

8.2.2 Enzyme preparation

Recombinant KDG-aldolase was purified as described in chapter 6. Elution from the final anion exchange column was in 20 mM Tris-HCl, pH 8.5, containing approximately 200 mM NaCl. Salt was removed by gel filtration in 20 mM Tris-HCl, pH 8.5, using a Superdex 200 column. The enzyme containing fractions were pooled and concentrated to 12 mg/ml by centrifugation at 1,500 g through Centricon 30 concentrators.

8.2.3 Crystallisation trials

Crystallisation trials were set up using the hanging drop diffusion method (Figure 8.1) in 24-well Linbro plates. The crystallisation reagents that were used are Crystal Screen (50 trials), PEG 6000 Grid Screen (24 trials) and ammonium sulphate Grid Screen (24 trials). 2 µl of KDG-aldolase (12 mg/ml) was mixed with an equal volume of crystallisation reagent and suspended, hanging from a siliconised cover slip, over a well containing 1 ml of reagent. Plates remained at room temperature, and were examined regularly for crystal growth.

8.2.4 Data collection

In house data collection was kindly performed by E. Hendry and G. Taylor at 293 K on a Cu Enraf-Nonius rotating-anode X-ray source (operating at 45 kV, 80mA) and using a 150 mm radius MAR image-plate detector.

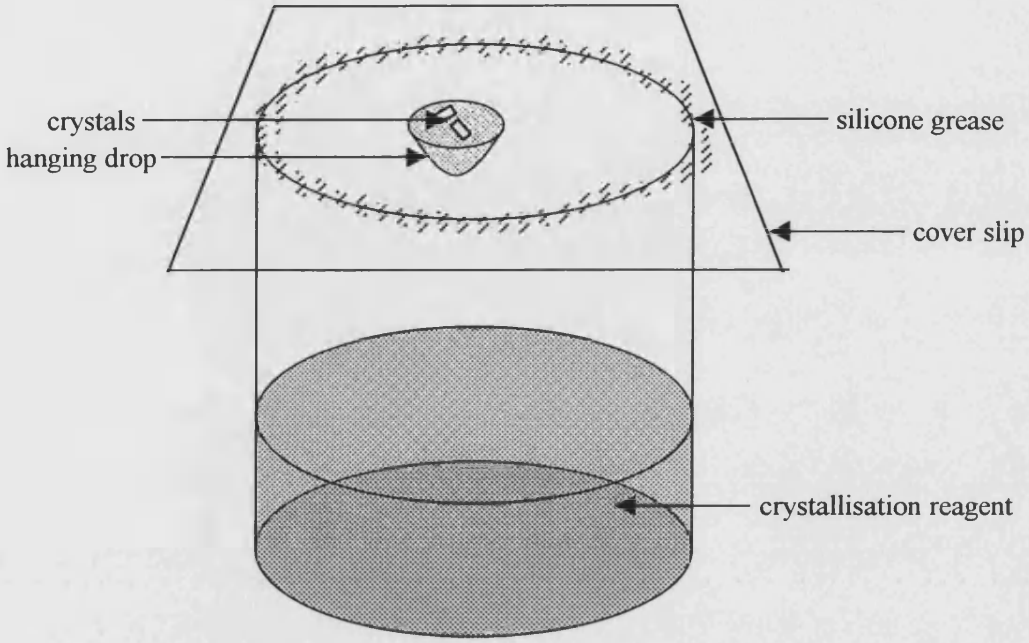


Figure 8.1 Growing crystals by the hanging drop diffusion method.

Synchrotron data were collected at the DESY synchrotron source in Hamburg by G. Taylor, G. Bell and E. Hendry at either beam line X31 (λ = 1.04 Å) or BW7b (λ = 0.8345 Å) at 100 K. 28 % (v/v) glycerol was used as a cryoprotectant

8.2.5 Data processing

The programmes DENZO and SCALEPACK were used to process diffraction data (Otwinowski and Minor, 1996). This section of work was performed by E. Hendry.

8.3 RESULTS

8.3.1 Crystal formation

Within 24 h crystals had appeared in seven different conditions, with three distinct morphologies. Table 8.1 summarises the formation of diffracting crystals over 9 days and Table 8.2 gives the composition of the reagents that gave crystals in this time period. Figure 8.2 shows pictures of the hexagonal (Crystal Screen, # 39) and orthorhombic crystals (Ammonium sulphate Grid Screen, # B1).

8.3.2 Data collection

The data for the hexagonal crystals (Crystal Screen, # 39) were collected in house, with diffraction to 5.0 Å. The unit cell dimensions were calculated to be a = b = 107 Å, c = 247 Å.

The orthorhombic data sets were collected at the DESY synchrotron source in Hamburg at 100 K using 28 % (v/v) glycerol as a cryoprotectant. The orthorhombic crystals (1.6 M ammonium sulphate, 0.1 M citric acid, pH 4) of space group $P2_12_12$ with unit cell dimensions a = 135 Å, b = 136 Å, c = 189 Å, diffracted to 2.16Å resolution.

The results for the diffraction conditions are summarised in Table 8.3.

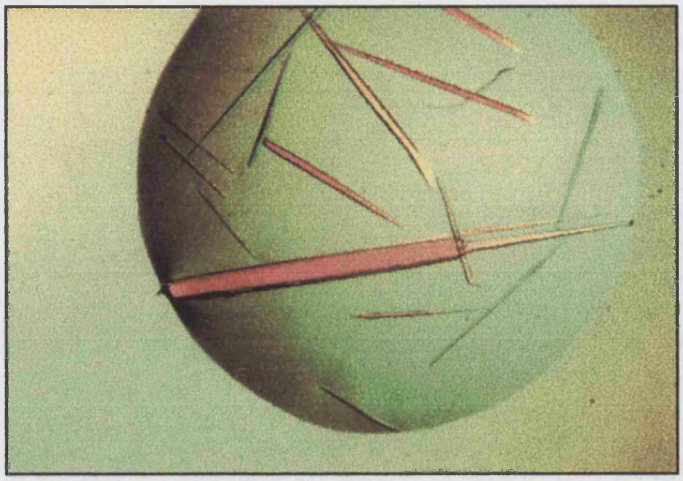
Data were processed for the orthorhombic form using DENZO and SCALEPACK (Otwinowski and Minor, 1996).

Table 8.1 Crystallisation trials of KDG-aldolase.

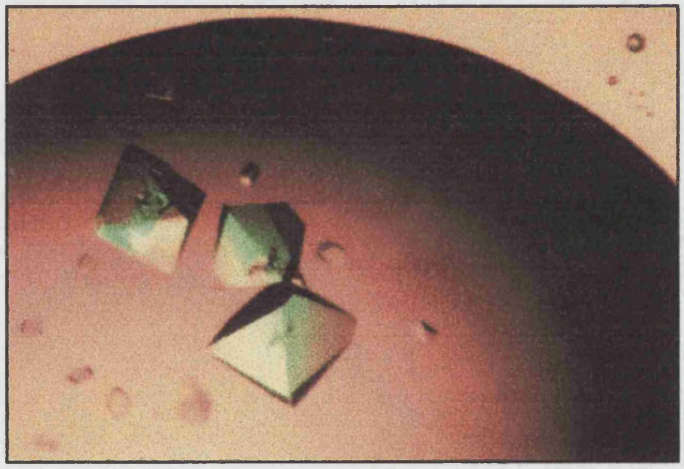
HAMPTON RESEARCH KIT	REF.	APPEARANCE			
		24 H	4 DAYS	5 DAYS	6 DAYS
CRYSTAL SCREEN	17			precipitate and needles	
	18			precipitate and needles	
	20	needles	needles	precipitate and needles	
	38			needles in stars	
	39				hexagonal rod crystals
	41	needles	needles	small needles	
	42	needles	needles	precipitate and needles	
	50	needles	needles	crystalline precipitate	
PEG 6000 GRID SCREEN	B1	needles	needles	bundles of sticks	
	B2			needles	
	B3	needles	needles, diffraction to 8Å	needles	
AMMONIUM SULPHATE GRID SCREEN	B1				orthorhombic crystals
	C1			orthorhombic crystals	
	C3			crystalline precipitate	

HAMPTON RESEARCH KIT REFERENCE	REAGENT COMPOSITION
CRYSTAL SCREEN	
17	30 % PEG 4000, 0.1 M Tris-HCl pH 8.5, 0.2 M lithium sulphate
18	20 % PEG 8000, 0.1 M Na Cacodylate pH 6.5, 0.2 M Magnesium acetate
20	25 % PEG 4000, 0.1 M Na Acetate pH 4.6, 0.2 M Ammonium sulphate
38	1.4 M Sodium citrate, 0.1 M Na Hepes pH 7.5
39	2 % PEG 400, 0.1 M Na Hepes pH 7.5, 2.0 M Ammonium sulphate
41	10 % iso-Propanol, 0.1 M Na Hepes pH 7.5, 20 % PEG 4000
42	20 % PEG 8000, 0.05 M Potassium phosphate
50	15 % PEG 8000, 0.5 M Lithium sulphate
PEG 6000 GRID SCREEN	
B1	10 % PEG 6000, 0.1 M Citric acid pH 4
B2	10 % PEG 6000, 0.1 M Citric acid pH 5
B3	10 % PEG 6000, 0.1 M MES pH 6
AMMONIUM SULPHATE GRID SCREEN	
B1	1.6 M Ammonium sulphate, 0.1 M Citric acid pH 4
C1	2.4 M Ammonium sulphate, 0.1 M Citric acid pH 4
C3	2.4 M Ammonium sulphate, 0.1 M Citric acid pH 6

Table 8.2 Reagent composition for each condition that gave KDG-aldolase crystals.



(a)



(b)

Figure 8.2 Photographs of KDG-aldolase crystals. Hexagonal rod crystals (a) formed in Crystal Screen # 39 (2 % PEG 400, 0.1 M Na Hepes pH 7.5, 2.0 M Ammonium sulphate) and orthorhombic crystals (b) formed in Ammonium sulphate Grid Screen, # B1 (1.6 M Ammonium sulphate, 0.1 M Citric acid pH 4).

Table 8.3 Summary of data for 3 crystal morphologies.

CRYSTAL MORPHOLOGY	HAMPTON RESEARCH REF.	REAGENT COMPOSITION	DIFFRACTION RESOLUTION	UNIT CELL DIMENSIONS	SPACE GROUP
Hexagonal	Crystal Screen, #39	2 % PEG 400, 0.1 M Na Hepes pH 7.5, 2.0 M Ammonium sulphate	5.0 Å	a = b = 107 Å, c = 247 Å	not determined
Orthorhombic	A.S. Grid Screen, # B1	1.6 M Ammonium sulphate, 0.1 M Citric acid pH 4	2.16 Å	a = 135 Å, b = 136 Å, c = 189 Å	$P2_12_12$

8.4 DISCUSSION

Crystallisation studies on KDG-aldolase are underway, with particular attention to solving the three-dimensional structure using the data set for the orthorhombic crystal form. The initial result of obtaining 2.16Å resolution data has provided a good indicator that this data set is worth proceeding with. Molecular replacement studies with the *E. coli* NAL as a phasing model appears to be the most promising course of action. Further discussion of the similarity of KDG-aldolase at the tertiary and quaternary level with the *E. coli* NAL and DHDPS enzymes requires a solution to the KDG-aldolase data set to be found.

Of particular interest would be an analysis of the structural factors conferring thermostability to the *S. solfataricus* KDG-aldolase. For example, increased rigidity, and/or reduced flexibility, of individual subunits (shorter loops and reduction in cavities), and strengthening of the intersubunit contacts (increase in complementarity and ionic networks) have been observed in citrate synthase (Danson and Hough, 1998). Similar traits may be found between the KDG-aldolase enzyme and its mesophilic counterparts from the NAL superfamily.

A crystal structure of the KDG-aldolase with pyruvate immobilised in the active site via sodium borohydride reduction would also be of interest. This would aid identification of key active site residues, which may confer substrate specificity. A rational approach to site directed mutagenesis of these residues might facilitate alteration of the wild type enzyme's specificity, which would increase the range of biotechnological applications of the enzyme.

9 KDG BIOTRANSFORMATION

9.1 INTRODUCTION

There has been varied interest in the production of KDG, a chemical not available commercially, with references dating back to 1960 (Merrick & Roseman, 1960) and covering more than 16 approaches, of which a number have preparative scale potential. These approaches generally fall into two categories, chemical and enzymatic syntheses, both of which can be further subdivided according to the exact nature of the production. This chapter outlines some of these syntheses, and provides the culmination to this thesis by the experimental investigation into the effectiveness of *S. solfataricus* KDG-aldolase in providing a novel route to the enzymatic production of KDG.

Biological interest in KDG arises from its role both as an intermediate of the modified Entner-Doudoroff glycolytic pathway (Entner and Doudoroff, 1952; De Rosa et al., 1984) and also as an intermediate of hexuronate metabolism in the phytopathogenic bacterium *Erwinia chrysanthemi* (Starr and Chatterjee, 1972). KDG, oringinating from galacturonte or glucuronate catabolism, has been shown to be the inducer of pectate-lyase synthesis by binding to a specific site on the KdgR protein, a repressor of the expression of the genes of the *kdg* regulon (Condemine et al., 1986; Hugouvieux-Cotte-Pattat & Robert-Baudouy, 1987). Pectinolysis, giving rise to soft rot disease of plants, results in the death of tissue, causing economically important damage to plants, either in the field or in storage after harvest.

Chemical interest in KDG appears to have arisen as a result of it not being commercially available, leading to the design of various synthetic pathways for the production of this elusive molecule. The complexity encountered in synthesis, whether by enzymatic or chemical procedures, the difficulty in recrystallisation of the product, and the challenge of

interpretation of analyses of a compound that exists in so many forms appears to have triggered more research into novel, simplified procedures.

KDG has also potential as a useful chemical building block, due to the four different oxidation states of five of its six carbon atoms. This allows possible chemical modification of these carbon atoms without lengthy protection and deprotection procedures, giving rise to novel compounds with new prospects.

9.1.1 Chemical syntheses of KDG

9.1.1.1 Portsmouth (1968)

Synthesis of KDG was by condensation of oxaloacetic acid and D-glyceraldehyde under neutral, aqueous conditions, followed by fractionation on Dowex-1 (formate) resin. Two epimers were produced in the ratio 10:1, the major one of which it was possible to isolate, crystallise and identify as 3-deoxy-D-*erythro*-hexulosonate. The second epimer was therefore concluded to be 3-deoxy-L-*threo*-hexulosonate. Figure 9.1 illustrates the mechanism involving nucleophilic attack on the most favoured rotamer of D-glyceraldehyde. Nucleophilic addition of the carbanion (pyruvate, resulting from the decarboxylation of oxaloacetic acid) to D-glyceraldehyde would be preferable from the least sterically hindered face, shown as mode II. The hydroxymethyl group of C3 hinders attack by mode I, as demonstrated practically by the 10:1 ratio of *erythro:threo* forms of KDG produced.

9.1.1.2 Chemical synthesis via Wittig-reagent

Two groups synthesised KDG via very similar routes, both starting with D-erythrose derivatives and involving Wittig reactions with either 2,4-*O*-isopropylidene-D-erythrose (Plantier-Royon et al., 1991), or 2,4-*O*-ethylidene-D-erythrose (Ramage et al., 1991). In the case of Plantier-Royon's group, the overall yield was 35%, in which 4 forms of KDG existing in equilibrium were detected by nmr: α-pyranose 10 %, β-pyranose 50 %, α-furanose 20 % and β-furanose 20 %. Ramage et al. (1991) crystallised KDG as the Ba^{2+} salt monohydrate, which had analytical data consistent with the α-pyranose form of KDG.

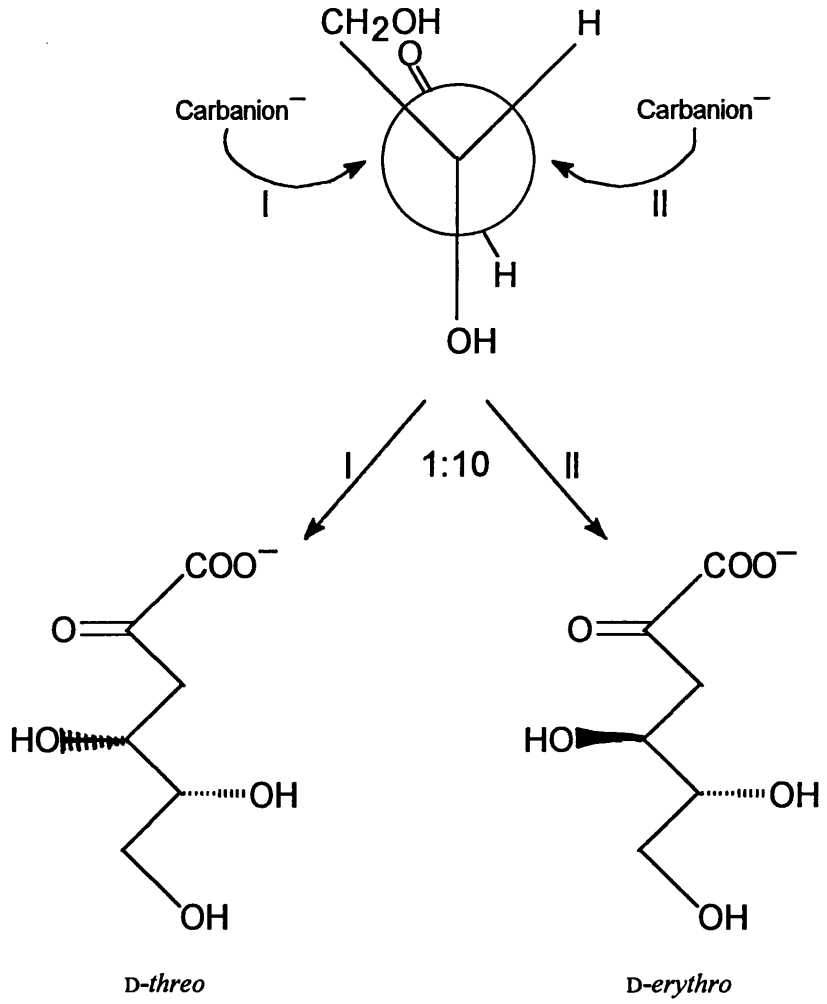


Figure 9.1 Mechanism of carbanion attack of D-glyceraldehyde (Portsmouth, 1968). The majority of synthesis occurs via mode II, giving rise to the D-*erythro* product. Mode I is sterically hindered by the presence of the hydroxymethyl group on C3 such that only a tenth of the product is of the D-*threo* form. In this instance, oxaloacetic acid was decarboxylated to give the carbanion pyruvic acid, CH_3-CO-CO_2^-.

9.1.1.3 Shimizu et al. (1996)

Shimizu et al. (1996) synthesised D-KDG via an eleven-step reaction which gives at least 90 % enantioselectivity. A mixture of both anomeric forms of pyranose and furanose structures of KDG was finally purified by IR $120H^+$ chromatography. The final yield was not given.

9.1.1.4 García et al. (1998)

Using a similar starting product to Ramage et al. (1991; see section 9.1.1.2 above), García et al. (1998) synthesised KDG. The condensation of monosaccharide derivatives with various stabilised diazo carbonyl compounds led to the production of β-acetoxy-α-diazo carbonyl compounds. A four-step reaction starting with 2,4-*O*-ethylidene-D-erythrose gave 92 mg KDG as the barium salt, representing a 37 % yield from the D-erythrose starting compound.

9.1.2 Enzymatic syntheses of KDG

Biotransformation approaches to the production of KDG have the advantage over chemical syntheses of higher degrees of stereoselectivity, reproducibility and yield (Sugai et al., 1993). The various routes taken to produce KDG by enzymatic syntheses can be classified into groups according to the enzyme(s) used.

9.1.2.1 KDG production by enzymatic dehydration of D-gluconate

This method has been carried out with both soluble cell extracts of *Alcaligenes* species M250 (De Ley et al., 1970; Kersters and De Ley, 1975) and *Clostridium pasteurianum* (Bender and Gottschalk, 1974), and also immobilised cells of *Sulfolobus solfataricus* (Nicolaus et al., 1986).

Kersters and De Ley (1975) report a 90 % conversion of D-gluconate to KDG after 15-20 h, followed by a 70 % yield after purification of the product on a Dowex-1 formate X8,

200-400 mesh column and crystallisation as the potassium salt. Enzyme was removed by perchloric acid precipitation.

Similarly, Bender and Gottschalk (1974) obtained a 99.9 % conversion of D-gluconate after a 30 min incubation at 50°C with a heat tolerant dehydratase, of which there was 40 % recovery following purification on Dowex 50 (H^+; 20-50 mesh) and crystallisation. Enzyme was recovered by ultrafiltration.

In the case of Nicolaus' group (Nicolaus et al., 1986), whole *S. solfataricus* cells were immobilised, facilitating re-use and avoiding labour intensive enzyme purification. Further breakdown of the product by KDG aldolase was prevented by inactivation with sodium borohydride. The reaction was carried out by passing D-gluconate through a column of immobilised cells with 65 % conversion to KDG, and the product recovered by purification on a silica gel column. The authors do not report the yield of the final product.

9.1.2.2 KDG synthesis by enzymatic dehydration of D-glucosaminic acid

Merrick and Roseman produced D-KDG from D-glucosaminic acid using cell extract (1960) and purified glucosaminic dehydrase (1966) from an unidentified organism, designated NRRL-P-826, which was isolated in the authors' laboratory by an enrichment culture technique. Approximately 91 % conversion of glucosaminic acid to KDG was obtained. The KDG was purified by anion exchange chromatography on Dowex 50, H^+ resin, and the eluate neutralised by the addition of $Ca(OH)_2$, giving 94 % isolation of the product as a crude calcium salt. Crystallisation was noted to be difficult in the presence of excess water, although possible by addition of a minimum amount of water to which several drops of acetone had been were added, followed by drying in a vacuum over P_2O_5.

9.1.2.3 KDG synthesis by enzymatic aldol condensation

Various filamentous fungi produce the enzyme KDG-aldolase when grown on glucose as the sole carbon source. Incubation of pyruvate and D-glyceraldehyde with resting cells of *Aspergillus niger* mycellium gives rise to a 65 % yield of a diastereomeric mixtue in a 1:24

ratio (Augé and Delest, 1993). NMR spectroscopy identified these products as 3-deoxy-D-*threo*-2-hexulosonate and 3-deoxy-D-*erythro*-2-hexulosonate (KDG) respectively, showing that there were two distinct aldolases acting with complementary facial selectivity. KDG was purified by anion exchange chromatography on AG 1-X8 (HCO_3^-, 100-200 mesh) resin, eluted with 0-0.4M ammonium bicarbonate, and isolated as the ammonium salt. Once again, the final yield and purity were not stated.

Whilst investigating substrate specificity of 3-deoxy-D-*manno*-2-octulosonic acid (KDO) aldolase from *Aureobacterium barkerei* strain KDO-37-2 (ATCC 49977), Sugai et al. (1993) found that both L- and D-glyceraldehyde were substrates for the enzyme. Using pyruvate as the natural donor gave L- and D-KDG respectively with rates 36 % and 23 % relative to the natural substrates. Although the efficiency of this enzyme in a larger scale biotransformation for the production of KDG has not been investigated, the authors report a 67 % yield of the natural product, KDO, from a multi-mmol scale reaction, with a 37 % overall yield of KDO as the ammonium salt monohydrate.

Similarly, Toone and co-workers (Allen *et. al.*, 1992; Shelton et. al., 1996) have investigated the ability of non-substrate specific KDPG aldolases to produce KDG via aldol condensation of D-glyceraldehyde and pyruvate. Preparative scale reactions with the enzymes from *Pseudomonas putida*, *Escherichia coli* and *Zymomonas molbilis* were carried out over a number of days with conversions of 25, 24 and 46 % respectively. Product was purified by anion exchange chromatography (AG-1X8, formate (HCO_2^-) form) with elution by formic acid, 0 – 1 M. Yields of the sodium salt in the mg range were 24 % *Pseudomonas putida*, 14 % *Escherichia coli* and 37 % *Zymomonas molbilis*.

9.1.2.4 KDG synthesis by enzymatic oxidation of 3-deoxyglucose

Having previously synthesised 3-deoxyglucose (3-deoxy-D-*erythro*-hexos-2-ulose), a one-step conversion by oxidation with D-glucose oxidase from *Aspergillus niger* resulted in a

65 % yield (based on consumed 3-deoxyglucose) of KDG (Madson and Feather, 1983). Pure product was obtained following ion exchange chromatography.

9.1.2.5 Two enzyme synthesis of KDG via 2-keto-3-deoxy-phospho-gluconate.

An alternative route to the production of KDG is via the production of KDPG, followed by an enzymatic dephosphorylation. O'Connell and Meloche (1982) report the conversion of 6-phosphogluconate to KDPG, using the enzyme 6-phosphogluconate dehydratase from glucose-grown *Pseudomonas putida*. They obtained a 79 % yield of 90 % pure KDPG. Although not performed by O'Connell and Meloche in this paper, Merrick and Roseman (1960) have obtained KDG from previously synthesised KDPG treated with sweet potato acid phosphatase. A combination of these two techniques gives the strategy often employed by Shelton et al. (personal communication) for the production of KDG.

The aim of the work reported in this chapter was to perform a large-scale KDG-aldolase biotransformation for the production of KDG and subsequently to purify this product from the enzyme and any remaining substrates. Analyses of the KDG for purity, conformation and configuration are also reported.

9.2 MATERIALS AND METHODS

9.2.1 KDG Biotransformation

0.5 M samples of DL-glyceraldehyde and Na-pyruvate were prepared in H_2O and their pH measured. DL-glyceraldehyde had a pH of 4 and was adjusted to pH 6.1 with NaOH. Na-pyruvate had a pH of 6.1 and was left unchanged. 50 ml of each substrate was added to 391 ml dH_2O at 70°C in a water bath. When the temperature had re-equilibrated to 70°C a time = 0 sample was withdrawn, and 9 ml of KDG-aldolase in 20 mM Tris, pH 8.5, 0.2 M NaCl, containing 232 Units in 22.5 mg, was added. The reaction was maintained at 70°C, and was continuously stirred from overhead. 0.5 ml samples were withdrawn every 30 min, the enzyme reaction stopped by addition to 0.5 ml 8 mM H_2SO_4, and the products analysed by RI-HPLC (see section 9.2.3). Further DL-glyceraldehyde was added at 1 h 20 min (2.7 g) and 4 h 15 min (0.9 g) and Na-pyruvate at 2 h 40 min (2.75 g). The pH was monitored and remained between pH 6 and 7. The reaction was stopped at 7 h by cooling to 4°C, and was stored at this temperature until further required. No significant change in substrate / product concentration was detected after 84 h at 4°C.

9.2.2 KDG Purification

In a method similar to that used by Simpson et al. (1966) and Kersters and De Ley (1975), KDG was separated from substrates and enzyme by anion exchange chromatography using Dowex 1-X8 formate, 200-400 mesh.

500 g of resin was prepared by suspension in 2 L of 4 M HCl and then successive washing in water by decantation to remove the fines. The resin was transferred to a sintered glass funnel over a Buchner flask and washed with water using a slight vacuum to remove excess HCl. Washings were monitored by measuring the pH, stopping at pH 5.5. The resin was then transferred to a beaker containing 0.5 L of 3 M Na-formate in 1 M formic acid and stirred gently for 30 min to replace chloride (Cl^-) ions on the Dowex resin with formate ions ($HCOO^-$). 1 L of the same solution was then drawn through the resin in the sintered glass

funnel. The resin was then stirred in 50 % (v/v) formic acid for 1 h, followed by washing with water over a vacuum to remove the chloride ions and acid. The presence of excess chloride ions was tested for, by adding a drop of filtrate to silver nitrate solution. If a cloudy precipitate forms then chloride ions are still present and the resin requires further washing. The pH was monitored to assess the removal of acid.

A 5 cm diameter column was packed with the resin to a height of 33 cm and washed thoroughly with water. The total column volume was 648 ml. Having equilibrated the column in water, the 0.5 L biotransformation mixture was loaded on to the column and the flow-through discarded. The column was then washed with a further 705 ml of water and the washings collected and analysed by RI-HPLC for substrate / product elution. A 2 L gradient of 0 - 0.4 M formic acid was run through the column. The initial 560 ml was collected as one pool, then 11 ml fractions were collected and analysed by RI-HPLC. The 2 L gradient was followed by 1 L of 0.4 M formic acid, which was followed by 0.5 L of 0.5 M formic acid and 1 L of 0.6 M formic acid. 11 ml fractions were collected throughout and analysed by RI-HPLC.

Fractions identified by RI-HPLC to contain KDG were combined into 3 pools according to content (see results section 9.3.3) and treated according to a modification of the protocol used by Kersters and De Ley (1975).

For pool 1, formic acid was removed by repeated evaporations under reduced pressure, and the concentrate neutralised by addition of 5 % (w/v) KOH. This was further concentrated to a syrup *in vacuo* and solubilised with warm methanol containing 1 % (v/v) water. Crystals of the K^+ salt of KDG formed overnight at 4°C and were collected by filtration and then dried by lyophilisation. Further crystallisation in the filtrate was noted and this was also dried by lyophilisation.

In the case of pools 2 and 3, some of the formic acid was removed by repeated evaporations under reduced pressure, followed by lyophilisation of the concentrate to yield sticky solids.

9.2.3 Analysis by proton nmr

^{1}H nmr spectra were recorded using a Varian Mercury-400 MHz spectrometer. The sample was in D_2O, and analysis was performed at room temperature. This section of work was performed by R. Kinsman and G. Buchanan.

9.2.4 Optical rotation

20 mg of pool 1 was diluted in 2 ml distilled water to give a sample of approximate 1 % (w/v) concentration. 1.8 ml of this sample was loaded into a vessel of length 10 cm, and inserted into the optical rotation chamber. The observed rotation was treated according to the Equation given below (Equation 9.1) to give the specific rotation.

$$[\alpha]_D^t = \frac{(\theta x 100)}{(c x l)} \qquad \textbf{Equation 9.1}$$

Where θ is the measured shift, c is the concentration in g / 100 ml of solvent, and l is the length of the cell in dm. Measurements were made at room temperature.

Procedures used in this chapter that are not mentioned in this section will be found in chapters 2 and 3.

9.3 RESULTS[1]

9.3.1 RI-HPLC calibration

Elution of compounds from ion exchange HPLC occurs at times specific to the chemical nature of the compound loaded, such that it is possible to assign retention times for single compounds to enable identification of compounds in mixtures. Table 9.1 gives retention times for DL-glyceraldehyde, Na-pyruvate, KDG and formic acid.

Compound	Retention time (min)
DL-Glyceraldehyde	13.7
Na-pyruvate	12.2
KDG	10.5 & 11.1
Formic acid	18.0

Table 9.1 HPLC retention times for DL-glyceraldehyde, Na-pyruvate, KDG and formic acid.

The RI-HPLC output gives a peak area proportional to the amount of compound injected. Having injected known amounts of DL-glyceraldehyde, Na-pyruvate, KDG and formic acid, calibration curves of peak area versus amount of chemical were plotted (Figures 3.7 and 9.2). The equations of the lines (Equations 3.4 – 3.6 and 9.2) can be used to calculate unknown amounts of these chemicals.

9.3.2 Reaction progress

The 0.5L KDG-aldolase biotransformation was followed by removing 0.5ml samples every 30 min, stopping by 1:1 dilution with 8mM H_2SO_4, and analysis by RI-HPLC. Figure 9.3 shows the RI-HPLC traces for samples at 30 min and 7 h. Peaks for each substrate and

[1] For the reader's convenience some details given in this section are a repeat of those given in chapter 3.

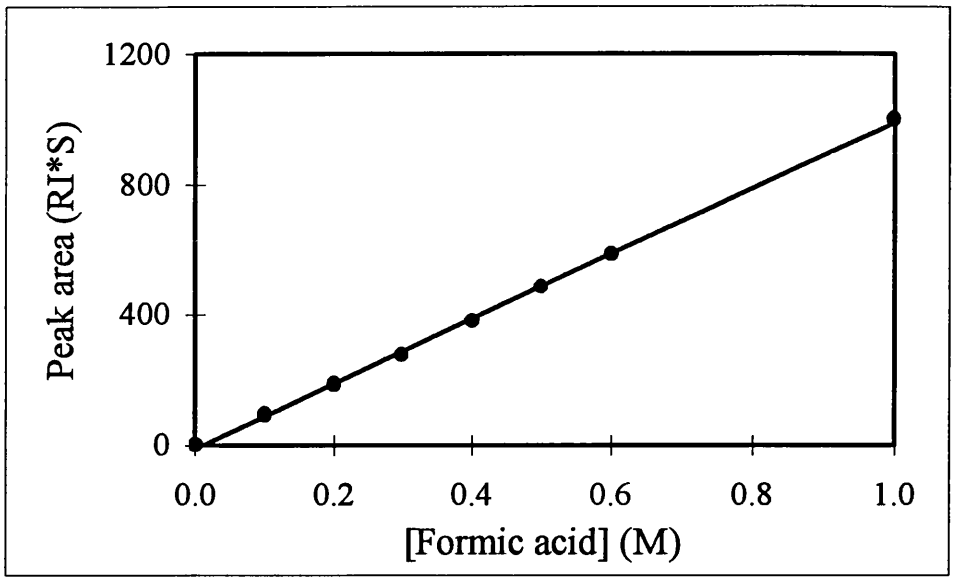
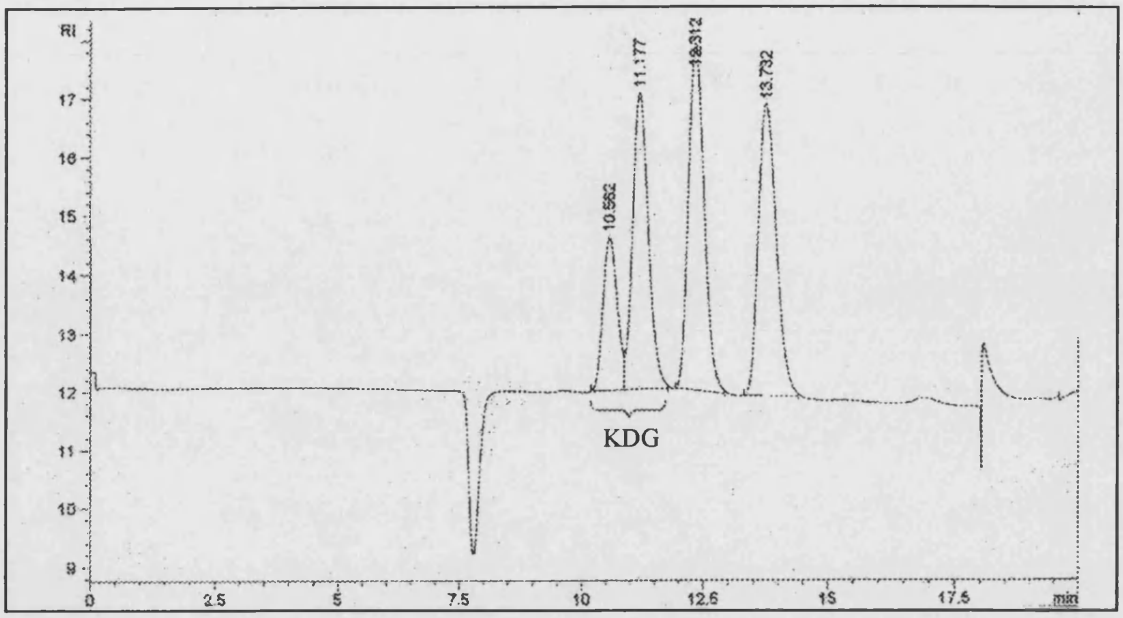
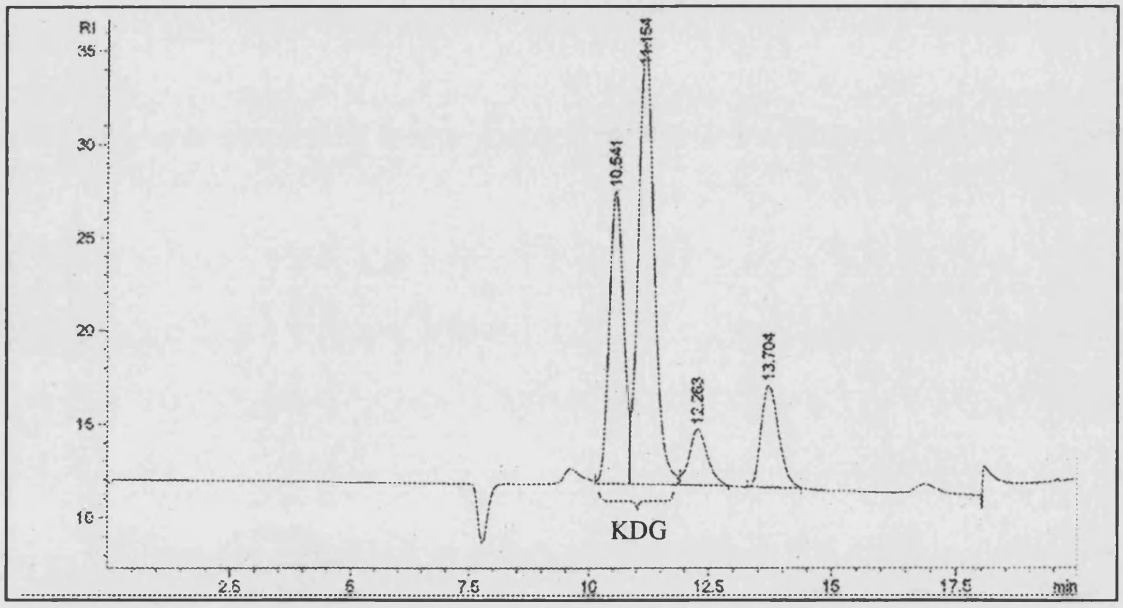


Figure 9.2 Calibration curve of RI-HPLC peak area against concentration (M) of formic acid in sample. The formic acid concentration is relative to the concentration of the sample *before* 1:1 dilution in H_2SO_4 mobile phase.

[Formic acid] (M) = (peak area + 12.264)/1002.3 **Equation 9.2**



(a) 30 min



(b) 7 h

Figure 9.3 RI-HPLC analysis of biotransformation progress. The amount of substrates, pyruvate (12.3 min) and DL-glyceraldehyde (13.7 min) can be seen to decrease as the two KDG product peaks (10.5 and 11.1 min) increase over time. Note that the y-axis scale for (b) is approximately double that of (a).

product are seen, with the area of the peak proportional to the amount of substrate or product in the sample injected. Using the Equations 3.4 – 3.6 and 9.2, the progress of the reaction was followed by calculating the amounts of substrate and product at each time point. Table 9.2 shows the results for each substrate and product and Figure 9.4 illustrates these Figures in graphical form.

Interestingly, the two product peaks that represent KDG formed during the reaction have an area ratio of 1.8 (peak 2 : peak 1). This is equal to the ratio of the K_m values for D-glyceraldehyde compared to L-glyceraldehyde.

The reaction was supplemented with additional Na-pyruvate and DL-glyceraldehyde over the course of the biotransformation, both to maintain the production of KDG and also to replenish DL-glyceraldehyde, which is unstable at 70°C over long periods.

The reaction was terminated at 7 h by cooling to 4°C. At this time there had been 7.6 g (0.04 moles) of KDG produced, and there remained 1.2 g (0.013 moles) DL-glyceraldehyde and 1.1 g (0.01 moles) Na-pyruvate. 5.85 g (0.065 moles) DL-glyceraldehyde and 5.5 g (0.05 moles) Na-pyruvate had been added in total, giving a conversion of 94 % relative to the Na-pyruvate reacted.

It should be noted that DL-glyceraldehyde undergoes thermal decomposition at 70°C, and hence the molar decrease of DL-glyceraldehyde concentration is larger than the molar increase in KDG concentration.

9.3.3 KDG Purification

The KDG biotransformation was loaded onto an anion exchange resin. The column was washed with water, during which time unreacted DL-glyceraldehyde was recovered. Following elution with the formic acid gradient, fractions 332-430 were shown by RI-HPLC to contain two KDG peaks as partially overlapping zones (Figure 9.5). Fractions 332-351 were pooled and named 'pool 1'. Fractions 352-405, containing the overlap between the two peaks, were pooled and named 'pool 2'. Fractions 406-430 were pooled, and named 'pool 3'.

product are seen, with the area of the peak proportional to the amount of substrate or product in the sample injected. Using the Equations 3.4 – 3.6 and 9.2, the progress of the reaction was followed by calculating the amounts of substrate and product at each time point. Table 9.2 shows the results for each substrate and product and Figure 9.4 illustrates these Figures in graphical form.

Interestingly, the two product peaks that represent KDG formed during the reaction have an area ratio of 1.8 (peak 2 : peak 1). This is equal to the ratio of the K_m values for L-glyceraldehyde compared to D-glyceraldehyde.

The reaction was supplemented with additional Na-pyruvate and DL-glyceraldehyde over the course of the biotransformation, both to maintain the production of KDG and also to replenish DL-glyceraldehyde, which is unstable at 70°C over long periods.

The reaction was terminated at 7 h by cooling to 4°C. At this time there had been 7.6 g (0.04 moles) of KDG produced, and there remained 1.2 g (0.013 moles) DL-glyceraldehyde and 1.1 g (0.01 moles) Na-pyruvate. 5.85 g (0.065 moles) DL-glyceraldehyde and 5.5 g (0.05 moles) Na-pyruvate had been added in total, giving a conversion of 94 % relative to the Na-pyruvate reacted.

It should be noted that DL-glyceraldehyde undergoes thermal decomposition at 70°C, and hence the molar decrease of DL-glyceraldehyde concentration is larger than the molar increase in KDG concentration.

9.3.3 KDG Purification

The KDG biotransformation was loaded onto an anion exchange resin. The column was washed with water, during which time unreacted DL-glyceraldehyde was recovered. Following elution with the formic acid gradient, fractions 332-430 were shown by RI-HPLC to contain two KDG peaks as partially overlapping zones (Figure 9.5). Fractions 332-351 were pooled and named 'pool 1'. Fractions 352-405, containing the overlap between the two peaks, were pooled and named 'pool 2'. Fractions 406-430 were pooled, and named 'pool 3'.

TIME	DL-GLYCERALDEHYDE		PYRUVATE		KDG	
(h)	(g)	(mmoles)	(g)	(mmoles)	(g)	(mmoles)
0	1.95	21.7	2.78	30.9	0.00	0.0
0.5	1.05	11.7	1.75	19.4	1.45	7.3
1	0.55	6.1	1.16	12.9	2.33	11.6
1.5	2.78	30.9	0.83	9.2	2.83	14.2
2	2.29	25.4	0.03	0.3	3.57	17.8
2.5	2.15	23.9	0.03	0.3	3.80	19.0
3	1.81	20.1	2.73	30.3	4.23	21.1
3.5	1.43	15.8	2.32	25.8	5.03	25.1
4	1.11	12.3	1.96	21.7	5.56	27.8
4.5	1.69	18.8	1.66	18.4	6.01	30.1
5	1.43	15.9	1.35	15.0	6.41	32.1
5.5	1.25	13.9	1.19	13.2	6.66	33.3
6	1.15	12.7	1.05	11.7	6.87	34.3
7	1.15	12.8	1.08	11.9	7.60	38.0

Table 9.2 Progress of biotransformation. As the amounts of the substrates decreases, there is a concomitant increase in KDG production. Additional DL-glyceraldehyde is added at 1 h 20 min and 4 h 15 min. Additional Na-pyruvate is added at 2 h 40 min.

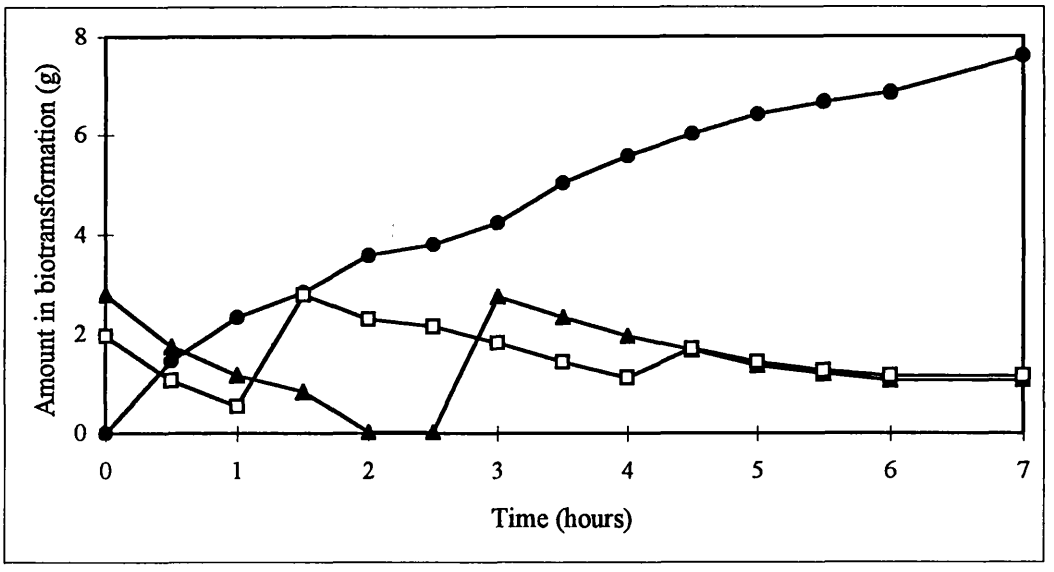


Figure 9.4 Graph illustrating biotransformation progress. The production of KDG (●) mirrors the consumption of DL-glyceraldehyde (□) and Na-pyruvate (▲). Additional DL-glyceraldehyde was added at 1 h 20 min and 4 h 15 min. Additional Na-pyruvate was added at 2 h 40 min.

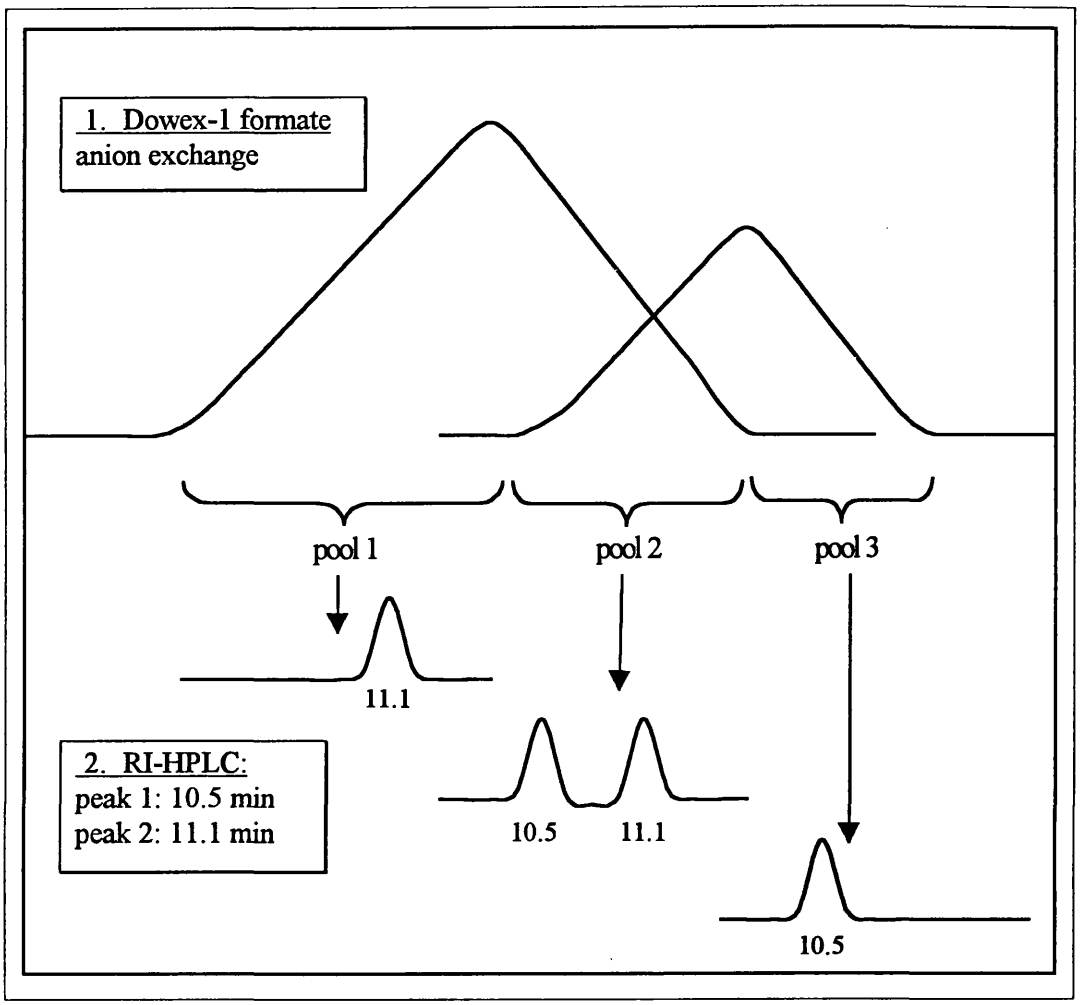


Figure 9.5 Diagrammatic representation of elution of two forms of KDG from Dowex anion exchange, and subsequent analysis of samples by RI-HPLC. Pool numbers are indicated, as are the elution times for peaks detected by RI-HPLC.

Pool number	Fractions	Volume (ml)	[Formic acid] (M)	KDG (g)	Yield (%)
1	332 – 351	175	0.444	1.92	25
2	352 – 405	530	0.524	3.8	50
3	406 – 430	240	0.596	0.1	1.3

Table 9.3 Data for each pool eluted from Dowex anion exchange. The amount of KDG purified is representative of the amount in g of dried solid, and is expressed as a percentage yield relative to the amount of pyruvate used in the biotransformation reaction.

The unreacted Na-pyruvate had not eluted by fraction 430. Table 9.3 gives details of amount of KDG recovered, and the concentration of formic acid in each pool.

Pool 1 was treated as per Kersters and De Ley (1975), giving 1.92 g of pale yellow powder of the K^+ salt of KDG, representing a yield of 25 % relative to the amount of KDG converted from substrates.

It was not possible to remove all the formic acid from pool 2 by repeated evaporations under reduced pressure, and so the material was freeze-dried, giving 3.8 g of a pale yellow, crystalline gum. This represents a yield of 50 % relative to the amount of KDG converted from substrates.

Pool 3 was similarly difficult to concentrate, and so was freeze-dried, giving 0.1 g of a sticky brown gum. This amounts to a yield of 1.3 % relative to the amount of KDG converted from substrates.

9.3.4 KDG analysis by RI-HPLC

By RI-HPLC the K^+ salt from pool 1 gives one clear peak, of retention time 11.3, equivalent to the second eluting peak on RI-HPLC from the biotransformation. Pools 2 and 3 were of greater complexity. They each had the expected peaks: 10.5 and 11.1 min for pool 2, and 10.5 min for pool 3, but in addition to this both displayed additional peaks around 8 and 9 min.

9.3.5 KDG analysis by ^{1}H nmr

The crystalline product from pool 1 was analysed by ^{1}H nmr[2]. Over 95 % of the nmr peaks were assigned to 4 species of KDG, all of them with the *erythro* configuration at the new stereocentre formed at C_4. The 4 species were identified as the α-pyranose, β-pyranose, α-furanose and β-furanose conformations of KDG, with molar ratios 10 % (α-pyranose), 52 % (β-pyranose), 16 % and 22 % (α- and β-furanoses). It was not possible to resolve whether

[2] ^{1}H nmr spectra are given in appendix 1.

the mixture contained both the D- and the L- isomers, or only one of the two.

9.3.6 KDG optical rotation

An optical rotation of pool 1 gave an observed rotation of +0.02°, which when treated according to Equation 8.3, gives a specific rotation of +2°. As the rotation is not zero, the sample cannot be an equal mixture of D- and L- isomers of KDG.

$$[\alpha]_D^t = \frac{(\theta x 100)}{(c x l)} \qquad \textbf{Equation 8.3}$$

Where, θ is the measured shift, c is the concentration is g / 100 ml of solvent, and l is the length of the cell in dm.

9.4 DISCUSSION

KDG is a chemical of multiple complexities, not least by way of the terminology used to describe it, or the many configurational isomers in which it can exist. Complications arise as chemists favour the 3-deoxy-D/L-*erythro/threo*-2-heulosonate terminology, whereas biochemists use the simpler term 2-keto-3-deoxygluconate, often abbreviated to KDG, which gives no indication of the configuration at either stereocentre.

Using precise nomenclature, KDG can exist as 3-deoxy-D-*erythro*-2-hexulosonate, 3-deoxy-D-*threo*-2-hexulosonate, 3-deoxy-L-*erythro*-2-hexulosonate, or 3-deoxy-L-*threo*-2-hexulosonate (Figure 9.6). To complicate matters further, these diastereomers can each exist as pyranoses or furanoses, each in either α or β anomeric forms, giving rise to 16 potential configurational isomers! It is also important to be aware that the hydroxyl group at the C_1 position would preferentially be axial, supporting the anomeric effect. This would cause the α-pyranose form of 3-deoxy-D-*threo*-hexulosonate to ring flip to the alternative conformation compared with the β-pyranose form, as shown in Figure 9.6. In the case of the α-pyranose of 3-deoxy-D-*erythro*-hexulosonate and the β-pyranoses of 3-deoxy-L-*erythro*- and *threo*-hexulosonate neither form is preferable, as in one case the hydroxyl group at the C_1 position would not be in the preferential axial position, and in the alternative conformation the hydroxyl groups on the other ring carbons would not be in the preferential equatorial position giving rise to steric hindrance. This effect would be marginally less pronounced in the case of 3-deoxy-L-*erythro*-hexulosonate as there would only be one unfavourable axial hydroxyl at the C_5 position when the C_1 hydroxyl is axial. In all three cases the isomer would probably exist in a conformation that is between each ring flip form (but *not* equilibrium between the two).

Having established the potential configurations and conformations that KDG *can* exist in, it is now possible to resolve which *is* produced by the biotransformation reaction between Na-pyruvate and DL-glyceraldehyde, catalysed by KDG-aldolase.

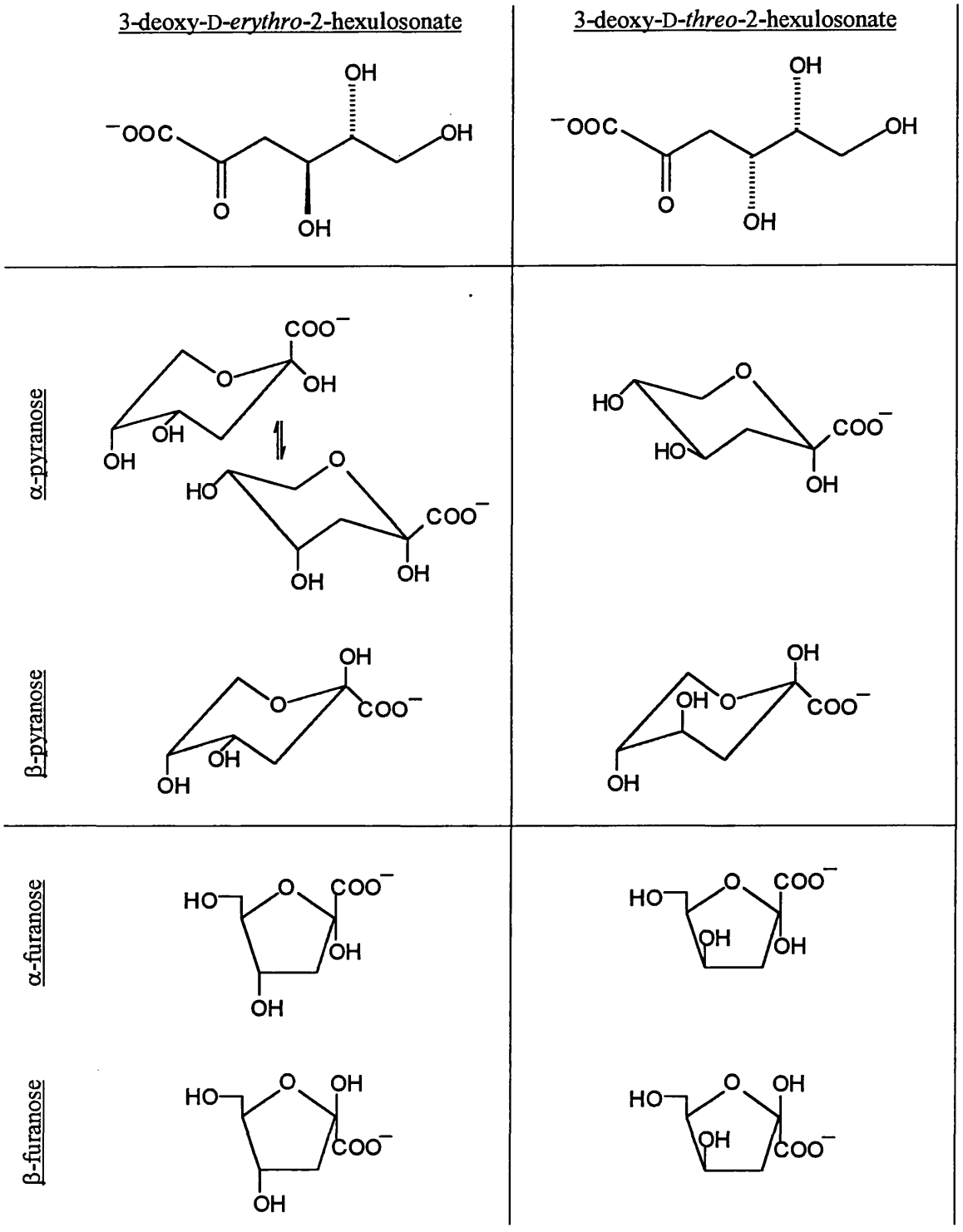
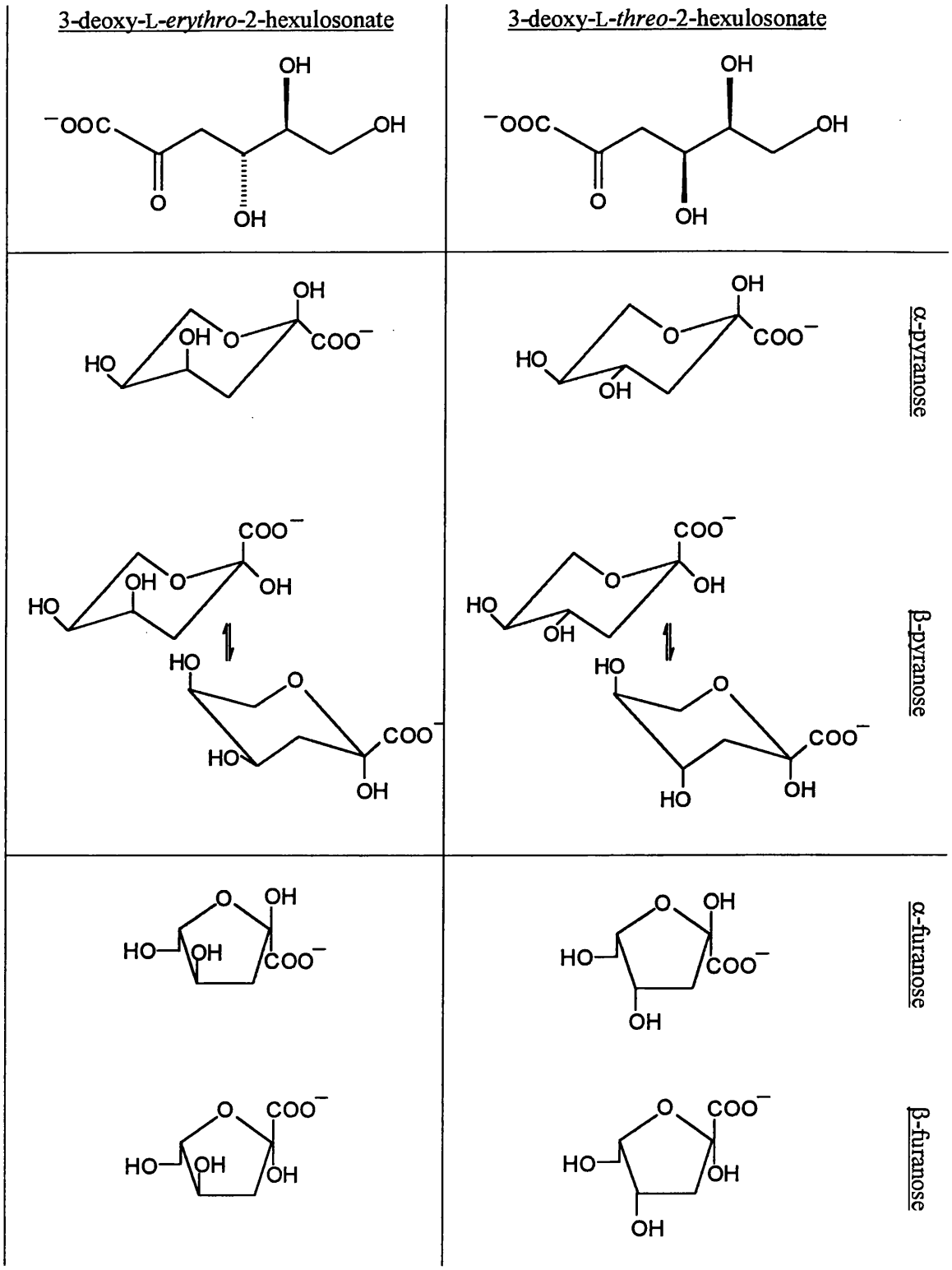


Figure 9.6 Configurational isomers of KDG. The top row illustrates the four, open chain configurational isomaers of KDG, labelled with their chemical names. Upon ring closure, a new chiral centre is formed at C_2, giving rise to α and β anomers of both pyranose (6 membered) and furanose (5 membered) rings. It is important to note that the α-pyranose of 3- deoxy-D-*erythro*-2-hexulosonate and the β-pyranose forms of both 3-deoxy-L-*erythro*-2-



hexulosonate and 3-deoxy-L-*threo*-2-hexulosonate are not stable in either chair conformation and probably exist as intermediates between the two possible conformations. 3-deoxy-D-*threo*-2-hexulosonate exists in the more stable chair conformation as illustrated.

Firstly, the results from RI-HPLC should be considered. As DL-glyceraldehyde is not resolved into two peaks, it can be assumed that D- and L-KDG would not be resolved on the basis of enantiomeric form at C_5 either. Therefore, the development of two product peaks for KDG must be due to another factor, such as the resolution of diastereomeric forms, i.e. *erythro / threo*. This raises the possibility of one peak representing L- and D-*erythro* forms of KDG and the other peak representing the L- and D-*threo* forms of KDG.

The second consideration is that of the NMR results for pool 1 (i.e. the second eluting peak detected by RI-HPLC), which was established to be the *erythro* form of KDG in different ring configurations, and possibly enantiomeric forms. This would suggest that pool 1 is potentially a mixture of L- and D-*erythro* forms of KDG and that pool 3 (i.e. the first eluting peak detected by RI-HPLC) is potentially a mixture of L- and D-*threo* forms of KDG, whereas pool 2 (the crossover of the two eluting peaks detected on RI-HPLC) is potentially all 4 forms. This is consistent with the first point made above.

Now consider the difference in K_m values for L- and D-glyceraldehyde in the KDG-aldolase reaction (section 7.3.2). The K_m for D-glyceraldehyde is 7.1 (± 0.6) mM, whereas that for L-glyceraldehyde is 3.9 (± 0.3) mM, the ratio being 1.8 (D:L). The V_{max} values, and therefore the relative rates (k_{cat}/K_m), of the stereoisomers when present as the racemate are the same. This suggests that KDG-aldolase would preferentially utilise D-glyceraldehyde, in turn producing 1.8 times as much D-KDG as L-KDG. Consideration of the areas of the two product peaks detected by RI-HPLC (section 8.3.2), shows them to have a ratio of 1.8 (peak 2 : peak1). This would suggest that, as detected by RI-HPLC, peak 2 (pool 1) represents L-KDG and peak 1 (pool 3) represents D-KDG.

Combining these three considerations, it is possible to rationalise that pool 1 (peak 2) must be the L-*erythro* form of KDG (3-deoxy-L-*erythro*-hexulosonate), and pool 3 (peak 1) the D-*threo* form (3-deoxy-D-*threo*-hexulosonate). Pool 2 is a mixture. Neither the L-*threo*,

or the D-*erythro* forms are produced in the biotransformation, thereby demonstrating that KDG-aldolase confers stereoselective C-C bond synthesis.

This is not an unusual observation for an aldolase enzyme. It has already been noted by Augé and Delest (1993) that the KDG-aldolase enzyme from *Aspergillus niger* catalyses the stereoselective bond formation between pyruvate and glyceraldehyde in a similar fashion. This result was somewhat confused, however, by the apparent contamination of the reaction mixture by a second aldolase with complementary facial selectivity. KDPG-aldolase likewise confers stereoselectivity, however the diasteromeric product has the *erythro* configuration relative to the D- enantiomeric substrate. The NAL enzyme, to which KDG-aldolase is probably similar in both structure and function, also catalyses stereoselective bond formation when operating with the natural substrate, *N*-acetyl-D-mannosamine (Fessner et al., 1992). In this case the acceptor carbonyl group is attacked from the *si* face (similar to mode II indicated in Figure 9.1), resulting in the formation of the *erythro* diastereomer (also described as *S* configuration). Interestingly, when presented with alternative aldehyde substrates, the attack is from the *re* face, resulting in formation of the *threo* diastereomer. DHDPS presents another example of stereoselectivity, giving rise to the *S* configuration of aldol condensation product (Blickling et al., 1997).

Following purification of the products, the overall yield resulting from the biotransformation described here was 76 %, and is represented by the D-*threo* and L-*erythro* forms of KDG. A significant fraction of the total amount converted to the L-*erythro* form (if the two forms are produced in the ratio 1.8, then a total yield of 64 % of L-*erythro* KDG would be anticipated) would be contained in pool 2, contaminated with the complementary diastereomer. However, if production of only one diastereomer were required, then optically pure glyceraldehyde could be used, thereby increasing pure product yield. In this instance, a 25 % yield of L-*erytho* KDG was purified as a light powder. It could be assumed that if, as observed here, 76 % of product was purified following a 94 % conversion, then a similar yield could be obtained of a single diastereomer, when using optically pure glyceraldehyde. This

results in a potential overall yield of 71 %. A comparison of this value with those given by the authors whose work was reviewed earlier in this chapter, indicates that this biotransformation is of notable potential for the large-scale production of diastereomerically pure KDG.

Of consequential value in any large-scale biotransformation of KDG would be the investigation of re-usability of the KDG-aldolase enzyme. On previous occasions enzymes have been successfully immobilised by attachment to inert supports such as Eupergit-C (production of *N*-acetyl-D-neuraminic acid using NAL enzyme [Mahmoudian et al., 1997]) or magnetic beads (glucose oxidase and glucoamylase for use in the sugar industry [Pieters and Bardeletti, 1992]), which following use can be easily retrieved and stored for re-use. Alternatively, enzyme has been recovered by ultrafiltration (Bender and Gottschalk, 1974). A second contributing factor to the re-usability of KDG-aldolase is the high degree of thermostability possessed by the enzyme.

In conclusion therefore, this chapter illustrates the potential of KDG-aldolase for stereoselective carbon-carbon bond synthesis. This biotransformation reaction proceeds favourably in the direction of production of KDG, and is followed by a simple purification from any remaining substrates. Application of this enzyme may be simply for KDG production, as KDG is not presently available commercially. KDG may have uses as a chemical precursor to a similar, more desirable target. Alternatively, further investigation of the substrate specificity of the enzyme, or even manipulation of the specificity, may allow the syntheses of more complex carbon based molecules with wide ranging applications.

10 GENERAL CONCLUSIONS AND FUTURE WORK

10.1 SUMMARY OF GENERAL CONCLUSIONS

This study has been successful in accomplishing the aim of purifying wild-type KDG-aldolase from *S. solfataricus*, cloning and sequencing the gene, expressing and purifying the recombinant enzyme from *E. coli*, and completing a general characterisation of the recombinant enzyme. To augment the enzyme characterisation, crystallographic studies of KDG-aldolase have been initialised, and a large-scale biotransformation reaction has been carried out to produce KDG. In conclusion, the main findings of this study can be summarised in the points below:

- At the primary sequence level, KDG-aldolase does not have any significant identity with other glycolytic aldolase enzymes, although the proposed overall quaternary fold of the enzyme (α/β_8 barrel) is related structurally to other aldolase enzymes.
- KDG-aldolase is proposed to be a member of the *N*-acetylneuraminate lyase superfamily of enzymes based on primary sequence analysis. This superfamily comprises NAL and DHDPS enzymes, as well as other dehydratases and decarboxylases. KDG-aldolase, NAL and DHDPS have all been shown to be homotetrameric enzymes.
- KDG-aldolase is a type I aldolase, forming a Schiff-base intermediate with either pyruvate or KDG via a lysine residue at the active site of the enzyme. The enzyme appears to be specific for the pyruvate donor, but variable for the glyceraldehyde acceptor. Certainly the enzyme is able to utilise both D- and L-glyceraldehyde.

- The enzyme is remarkably thermostable and has a high temperature of optimum indicating that this enzyme will be useful in biotransformations where enzyme stability is of paramount importance.
- KDG-aldolase has considerable potential for use in the large-scale bioproduction of enantiomerically pure KDG. The reaction and purification of product(s) are simple, and there is the potential for the process to be scaled-up.

10.2 SUGGESTIONS FOR FUTURE WORK

This study can be viewed as a neatly completed section of work; however, as with many fields of study it can likewise be viewed as a gateway to much further investigation. In the long-term, development of this enzyme for use in biotransformations requires an extension of the characterisation work already begun.

Clearly the crystallographic structure of KDG-aldolase requires more work for a solution to be reached. This is presently underway at the University of St. Andrews, Scotland, under the guidance of Prof. Garry Taylor. Having obtained a three-dimensional structure, the similarity to NAL, DHDPDS and other aldolases can be more accurately established. Predictions concerning the reaction mechanism can be made and computer-assisted modelling of the substrates in the active site will allow some degree of anticipation of the specificity of the enzyme. An eventual solution for the crystal structure of the enzyme with substrate(s) will provide more information about the arrangement of the substrate(s) in the active site.

Development of a suitable assay technique for the analysis of substrate specificity should be carried out, possibly concentrating on the RI-HPLC method. Ideally, on-line mass spectrometry would allow the rapid and accurate determination of product formation. Information about the parameters of substrate specificity would aid predictions concerning the active-site environment.

The recently developed technique of directed evolution by random mutagenesis may be used to create a library of mutants that can be screened for various altered attributes. Such benefits may be a decreased optimum temperature of activity, an increased range of substrate specificity, or even novel specificity for a particular substrate not already accepted by the enzyme.

The combination of knowledge gained from all the experimental work above could be assimilated in the rational design of KDG-aldolase mutants with specified properties desirable to industrial applications of the enzyme. It is certainly apparent that the use of enzymes in

industry is developing into a buyer's market, where the requirements for the reaction can be listed and an enzyme engineered to those specifications. Continuing work on KDG-aldolase will help to place this enzyme into the marketplace.

REFERENCES

Adams, M. W. W., Perler, F. B. and Kelly, R. M. (1995) Extremozymes: expanding the limits of biocatalysis. *Bio/Technology.* 13, 662-668.

Aguilar, C. F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M. and Pearl, L. H. (1997) Crystal structure of the β-glycosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*: resilience as a key factor in thermostability. *J. Mol. Biol.* 271, 789-802.

Aisaka, K., Igaraschi, A., Yamaguchi, K. and Uwajima, T. (1991) Purification, crystallisation and characterisation of *N*-acetylneuraminate lyase from *Escherichia coli*. *Biochem. J.* 276, 541-546.

Allen, S. T., Heintzelman, G. R. and Toone, E. J. (1992) Pyruvate aldolases as reagents for stereospecific aldol condensatation. *J. Org. Chem.* 57, 426-427.

Altschul, S. F., Gish. W., Miller. W., Myers. E. W. and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389-3402.

Amann, R. I., Ludwig, W. and Schleifer, K-H. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59, 143-167.

Augé, C. and Delest, V. (1993) Microbiological aldolisations. Synthesis of 2-keto-3-deoxy-D-gluconate. *Tetrahedron Assymetry.* 4, 1165-1168.

Bender, R. and Gottschalk, G. (1973) Purification and properties of D-gluconate dehydratase from *Clostridium pasteurianum*. *Eur. J. Biochem.* 40, 309-321.

Bender, R. and Gottschalk, G. (1974) Enzymatic synthesis of 2-keto-3-deoxy-D-gluconate from D-gluconate. *Anal. Biochem.* 61, 275-279.

Blickling, S., Renner, C., Laber, B., Pohlenz, H-D., Holak, T. A. and Huber, R. (1997) Reaction mechanism of *Escherichia coli* dihydrodipicolinate synthase investigated by X-ray crystallography and NMR spectroscopy. *Biochemistry.* 36, 24-33.

Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilising the principle of protein-dye binding. *Anal. Biochem.* 72, 284-255.

Brown, S. H., Sjoholm, C. and Kelly, R. M. (1993) Purification and characterisation of a highly thermostable glucose isomerase produced by the extremely thermophilic eubacterium, *Thermotoga maritima. Biotechnol. Bioeng.* 41, 878-886.

Budgen, N. and Danson, M. J. (1986) Metabolism of glucose via a modified Entner-Doudoroff pathway in the thermoacidophilic archaebacterium *Thermoplasma acidophilum. FEBS.* 196, 207-210.

Chan, M. K., Mukund, S., Kletzin, A., Adams, M. W. W. and Rees, D. C. (1995) Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferrodoxin oxidoreductuctase. *Science.* 267, 1463-1469.

Cheetham, P. S. J. (1998) What makes a good biocatalyst. *J. Biotechnol.* 66, 3-10.

Chen, K. and Arnold, F. H. (1993) Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. USA.* 90, 5618-5622.

Choi, I-G., Cho, S. K. and Yu Y. G. (1998) Overproduction, purification and characterisation of heat stable aldolase from *Methanococcus jannaschii*, a hyperthermophilic archaea. *J. Biochem. Mol. Biol.* 31, 130-134.

Christensen, P. N., Thomsen, G. D. and Branner, S. (1987) Development of detergent enzymes. In *Proceedings of the 2nd World Conference on Detergents, Montreux, Switzerland.* (Baldwin, A. R., Ed.), pp 181-186.

Claes, P., Dunford, M., Kenney, A. and Vardy, P. (1992) On-line dynamic light scattering for macromolecular characterisation. In *Laser Light Scattering in Biochemistry.* The Royal Society of Chemistry, Cambridge, UK. (Harding, S.E., Sattelle, D. B. and Bloomfield, V. A., Ed.). pp.66-76.

Comb, D. G. and Roseman, S. (1960) The Sialic Acids. I. The Structure and Enzymatic synthesis of *N*-acetylneuraminic acid. *J. Biol. Chem.* 235, 2529-2537.

Condemine, G., Hugouvieux-Cotte-Pattat, N. and Robert-Baudouy, J. (1987) Isolation of *Erwinia chrysanthemi kduD* mutants altered in pectin degradation. *J. Bacteriol.* 165, 937-941.

Connaris, H., West, S. M., Hough, D. W. and Danson, M. J. (1998) Cloning and overexpression in *Escherichia coli* of the gene encoding citrate synthase from the hyperthermophilic Archaeon *Sulfolobus solfataricus. Extremophiles.* 2, 61-66.

Corfield, A. P. and Schauer, R. (1982) In *The Sialic Acids.* Springer. New York. (Schauer, R., Ed.) p 195.

Crameri, A., Raillard, S-A., Bermudez, E. and Stemmer, W. P. C. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature.* 391, 288-291.

Dalby, A., Dauter, Z. and Littlechild, J. A. (1999) Crystal structure of human muscle aldolase complexed with fructose 1,6-bisphosphate: mechanistic implications. *Prot. Sci.* 8, 291-297.

Daly, M. J., Ouyang, L. and Minton, K. W. (1994) *In vivo* damage and recA-dependent repair of plasmid and chromosomal DNA in the radioresistant bacterium *Deinococcus radiodurans. J. Bacteriol.* 176, 3508-3517.

Danson, M. J. and Hough, D. W. (1998) Structure, function and stability of enzymes from the Archaea. *Trends Microbiol.* 6, 307-314.

D'Aquino, R. L. (1999) Biocatalysis makes headway in chemicals. *Chemical Engineering.* 106, 37-43.

De Ley, J., Kersters, K., Khan-Matsubara, J. and Shewan, J. M. (1970) Comparative D-gluconate metabilism and DNA base composition in *Archromobacter* and *Alcaligenes. Antonie van Leeuwenhoek.* 36, 193-207.

De Montigny C. and Sygusch, J. (1996) Functional characterisation of an extreme thermophilic class II fructose-1,6-bisphosphate aldolase. *Eur. J. Biochem.* 241, 243-248.

De Rosa, M., Gambacorta, A. and Bu' Lock, J. D. (1975) Extremely thermophilic acidophilic bacteria convergent with *Sulfolobus solfataricus. J. Gen. Microbiol.* 86, 156-164.)

De Rosa, M., Gambacorta, A., Nicolaus, B., Giardina, P., Poerio, E. and Buonocore, V. (1984) Glucose metabolism in the extreme thermoacidophilic arachaebacterium *Sulfolobus solfataricus. Biochem. J.* (1984) 224, 407-414.

De Vendittis, E. and Bocchini, V. (1996) Protein-encoding genes in the sulfothermophilic archaea *Sulfolobus* and *Pyrococcus. Gene.* 176, 27-33.

Dhar, N. M. and Altekar, W. (1986) Distribution of class I and class II fructose bisphosphate aldolases in halophilic archaebacteria. *FEMS Microbiol Lett.* 35, 177-181.

Dong, G., Vieille, C., Savchenko, A. and Zeikus, J. G. (1997) Cloning, sequencing, and expression of the gene encoding extracellular α-amylase from *Pyrococcus furiosus* and biochemical characterisation of the recombinant enzyme. *Appl. Environ. Microbiol.* 63, 3569-3576.

Dym, O., Mevarech, M. and Sussman, J. L. (1995) Structural features that stabilize halophilic malate dehydrogenase from an Archaebacterium. *Science.* 267, 1344-1346.

Eaton, R. W. (1994) Organisation and evolution of naphthalene catabolic pathways: sequence of the DNA encoding 2-hydroxychromene-2-carboxylate isomerase and *trans-o*-hydrobenzylidenepyruvate hydratase-aldolase from the NAH7 plasmid. *J. Bacteriol.* 176, 7757-7762.

Eisenthal, R. and Cornish-Bowden, A. (1974) The direct linear plot. A new graphical procedure for estimating enzyme kinetic parameters. *Biochem. J.* 139, 715-720.

Ensley, B. D. Ratzkin, B. J., Osslund, T. D., Simon, M. J., Wackett, L. P. and Gibson, D. T. (1983) Expression of napthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. *Science* 222, 167-169.

Entner, N. and Doudoroff, M. (1952) Glucose and gluconic acid oxidation by *Pseudomonas saccharophila. J. Biol. Chem.* 196, 853-862.

Farber, G. K. and Petsko, G. A. (1990) The evolution of α/β barrel enzymes. *Trends Biochem. Sci.* 15, 228-234.

Feinberg, A. P. and Vogelstein, B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132, 6-13.

Feinberg, A. P. and Vogelstein, B. (1984) Addition. *Anal. Biochem.* 137, 266-267.

Fessner, W-D., Badia, J., Eyrisch, O., Scheider, A. and Sinerius, G. (1992) Enzymes in organic synthesis. 5. Enzymic syntheses of rare ketose 1-phosphates. *Tetrahedron Lett.* 33, 5231-34.

Gamblin, S. J., Davies, G. J., Grimes, J. M., Jackson, R. M., Littlechild, J. A. and Watson, H. C. (1991) Activity and specificity of human aldolases. *J. Mol. Biol.* 219, 573-576.

García, F, S., Cebrián, G. M. P., López, A. H. and Herrera, F. J. L. (1998) Reactions of monosaccharide derivatives with diazocarbonyl compounds. Reactivity and synthetic applications. *Tetrahedron.* 54, 6867-6896.

Gerike, U., Danson, M. J., Russell, N. J. and Hough, D. W. (1997) Sequencing and expression of he gene encoding a cold-active citrate synthase from an Antartic bacterium, strain DS2-3R. *Eur. J. Biochem.* 248, 49-57.

Gottschalk, G. and Bender, R. (1982) D-Gluconate dehydratase from *Clostridium pasteurianum. Meth. Enzymol.* 90, 283-287.

Grogan, D. W. (1989) Phenotypic characterisation of the archaebacterial genus *Solfolobus*: comparison of five wild-type strains. *J. Bacteriol.* 171, 6710-6719.

Harris, E. L. V. and Angal, S. (1989) In *Protein purification methods, a practical approach.* IRL Press, OUP. p. 155.

Hensel, R., Jakob, I., Scheer, H. and Lottspeich, F. (1992) Proteins from hyperthermophilic archaea: stability towards covalent modification of the peptide chain. *Biochem. Soc. Symp.* 58, 127-133.

Higgins, D. G. and Sharp, P. M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene.* 73, 237-244.

Horecker, B. L., Tsolas, O. and Lai, C. Y. (1972) Aldolases. In *The Enzymes.* Academic Press, New York. (Boyer, P. D., Ed.), 3rd edition. Vol 7, pp.213-258,

Hough, D. W. and Danson, M. J. (1999) Extremozymes. *Curr. Op. Chem. Biol.* 3, 39-46.

Hugouvieux-Cotte-Pattat, N. and Robert-Baudouy, J. (1987) Hexuronate catabolism in *Erwinia chrysanthemi. J. Bacteriol.* 169, 1223-1231.

Hyun, H. H. and Zeikus, J. G. (1985) General biochemical characterisation of thermostable pullulanase and glucoamylase from *C. thermohydrosulfuricum. Appl. Environ. Microbiol.* 49, 1168-1173.

von Itstein, M., Wu, W-Y., Kok, G. B., Pegg, M. S., Dyason, J. C., Jin, B., Phan, T. V., Smythe, M. L., White, H. F., Oliver, S. W. et al. (1993) Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature.* 363, 418-423.

Izard, T., Lawrence, M. C., Malby, R. L., Lilley, G. G. and Colman, P. M. (1994) The three-dimensional structure of *N*-acetylneuraminate lyase from *Escherichia coli. Structure.* 2, 361-369.

Jeffcoat, R., Hassall, H. and Dagley, S. (1969a) The metabolism of D-glucarate by *Pseudomonas acidovorans. Biochem. J.* 115, 969-976.

Jeffcoat, R., Hassall, H. and Dagley, S. (1969b) Purification and properties of the D-4-deoxy-5-oxoglucarate hydro-lyase (decarboxylating). *Biochem. J.* 115, 977-983.

Johnson, M. L., Correia, J. J., Yphantis, D. A. and Halvorson, H. R. (1981) Analysis of data from the analytical ultracentrifuge by non-linear least-squares technique. *Biophys. J.* 36, 575-588.

Jurgens, G. And Saano, A. (1999) Diversity of soil Archaea in boreal forest before, and after clear-cutting and prescribed burning. *FEMS Microb. Ecol.* 29, 205-213.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. and Shore, V. C. (1960) Structure of myoglobin. A three-dimensional Fourier synthesis at 2 Å resolution. *Nature.* 185, 422-427.

Kersters, K. and De Ley, J. (1968) The occurrence of the Entner-Doudoroff pathway in bacteria. *Antonie van Leeuwenhoek.* 34, 393-408.

Kersters, K. and De Ley, J. (1975) 2-Keto-3-deoxy-D-gluconate. *Meth. Enzymol.* 41, 99-101.

Kersters, K., Khan-Matsubara, J., Nelen, L. and De Ley, J. (1971) Purification and properties and D-gluconate dehydratase from Achromobacter. *Antonie van Leeuwenhoek.* 37, 233-246.

Kim, B. S., Lee, S. C., Lee, S. Y., Chang, H. N., Chang, Y. K. and Woo, S. I. (1994) Production of poly(3-hydroxybutyric-co-3-hydroxyvaleric acid) by fed-batch culture of *Alcaligenes*

eutrophus with substrate control using on-line glucose analyser. *Enzyme Microb. Technol.* 16, 556-561.

Kim, M-J, Hennen, W. J., Sweers, M. and Wong C-H. (1988) Enzymes in carbohydrate synthesis: *N*-acetylneuraminic acid aldolase catalysed reactions and preparation of *N*-acetyl-2-deoxy-D-neuraminic acid derivatives. *J. Am. Chem. Soc.* 110, 6481-6486.

Kirino, H., Aoki, M., Aoshima, M., Hayashi, Y., Ohba, M., Yamagishi, A., Wakagi, T. and Oshima, T. (1994) Hydrophobic interaction at the subunit interface contributes to the thermostability of 3-isopropylmalate dehydrogenase from an extreme thermophile, *Thermus thermophilus. Eur. J. Biochem.* 220, 275-281.

Koch, R., Spreinat, A., Lemke, K. and Antranikian, G. (1991) Purification and properties of a hyperthermoactive α-amylase from archaebacterium *Pyroccocus woesei. Arch. Microbiol.* 155, 572-578.

Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946-950.

Kuchner, O. and Arnold, F. H. (1997) Directed evolution of enzyme catalysis. *Trends Biotechnol.* 15, 523-530.

Kurz, L. C., Roble, J. H., Nakra, T., Drysdale, G. R., Buzan, J. M., Schwartz, B. and Drueckhammer, D. G. (1997) Ability of single-site mutants of citrate synthase to catalyse proton transfer from the methyl group of dethiacetyl-coenzyme A, a non-thioester substrate analog. *Biochemistry.* 36, 3981-3990.

Laber, B., Gomis-Rüth, F-X., Romão M. J. and Huber, R. (1992) *Escherichia coli* dihydrodipicolinate synthase. Identification of the active site and crystallisation. *Biochem. J.* 288, 691-695.

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.* 227, 680-685.

Lange, C. C., Wackett, L. P., Minton, K. W. and Daly, M. J. (1998) Engineering a recombinant *Deinococcus radiodurans* for organopollutant degradation in radioactive mixed waste environments. *Nature Biotechnol.* 16, 929-933.

Langois, D. P and Dale, J. K. (1940) *US Patent* 2,202,609.

Lawrence, M. C., Barbosa, J. A. R. G., Smith, B. J., Hall, N. E., Pilling, P. A., Ooi, H. C. and Marcuccio, S. M. (1997) Structure and mechanism of a sub-family of enzymes related to *N*-acetylneuraminate lyase. *J. Mol. Biol.* 266, 381-399.

Little, J. W. and Mount, D. W. (1982) The SOS regulatory system of *Escherichia coli. Cell.* 29, 11-22.

Littlechild, J. A. and Watson, H. C. (1993) A data-based reaction mechanism for type I fructose bisphosphate aldolase. *Trends Biochem. Sci.* 18, 36-39.

Liu, J. Q., Odani, M., Dairi, T., Itoh, N., Shimizu, S. and Yamada, H. (1999) A new route to L-threo-3-[4-(methylthio)phenylserine], a key intermediate for the synthesis of antibiotics: recombinant low- specificity D-threonine aldolase-catalyzed stereospecific resolution. *Appl. Microbiol. Biotechnol.* 51, 586-591.

Lotz, B. T., Gasparski, C. M., Peterson, K. and Miller, M. J. (1990) Substrate specificity studies of aldolase enzymes for use in organic synthesis. *J. Chem. Soc. Chem. Commun.* 1107-1109.

MacGee, J. and Doudoroff, M. (1954) A new phosphorylated intermediate in glucose oxidation. *J. Biol. Chem.* 210, 617-626.

Madson, M. A. and Feather, M. S. (1983) The oxidation of 3-deoxy-D-*erythro*-hexos-2-ulose ("3-deoxyglucosone") to 3-deoxy-D-*erythro*-2-hexulosonic acid ("2-keto-3-deoxy-D-gluconate") by D-glucose oxidase. *Carbohydr. Res.* 115, 288-291.

Mahmoudian, M., Noble, D., Drake, C. S., Middleton, R. F., Montgomery, D. S., Piercey, J. E., Ramlakhan, D., Todd, M. and Dawson, M. J. (1997) An efficient process for production of *N*-acetyl-D-neuraminic acid using *N*-acetyl-D-neuraminic acid aldolase. *Enzyme and Microb. Technol.* 20, 393-400.

Matthews, B. W., Nicholson, H. and Becktell, W. J. (1987) Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc. Natl. Acad. Sci. USA.* 84, 6663-6667.

Mavridis, I. M., Hatada, M. H., Tulinsky, A. Lebioda, L. (1982) Structure of 2-keto-3-deoxy-6-phosphogluconate aldolase 2.8 Å resolution. *J. Mol. Biol.* 162, 419-444.

McCoy, M. (1999) Biocatalysis grows for drug synthesis. *Chemistry & Engineering.* Vol, 10-14.

McCutchen, C. M., Duffaud, G. D., Leduc, P., Peterson, A. R. H., Tayal, A., Khan, S. A. and Kelly, R. M. (1996) Characterisation of extremely thermostable enzymatic breakers (α-1,6-galactosidase and β-1,4-mannanase) from the hyperthermophilic bacterium *Thermotoga neapolitant* 5068 for hydrolysis of guar gum. *Biotechnol. Bioeng.* 52, 332-339.

Menéndez-Arias, L. and Argos, P. (1989) Engineering protein thermal stability. Sequence statistics point to residue substitutions in α-helices. *J. Mol. Biol.* 206, 397-406.

Merrick, J. M. and Roseman, S. (1960) Glucosamine metabolism. *J. Biol. Chem.* 235, 1274-1280.

Merrick, J. M. and Roseman, S. (1966) D-Glucosaminic acid dehydrase. *Meth. Enzymol.* 9, 657-660.

Milburn, D., Laskowski, R. and Thornton, J. (1998) Sequences annotated by structure: a tool to facilitate the use of structural information in sequence analysis. *Protein Engineering.* 11, 855-859.

Mirwaldt, C., Korndoerfer, I. and Huber, R. (1995) The crystal structure of dihydrodipicolinate synthase from *Escherichia coli* at 2.5 Å resolution. *J. Mol. Biol.* 246, 227-239.

Moore, J. C., Jin, H-M., Kuchner, O. and Arnold, F. (1997) Strategies for *in vitro* evolution of protein function: enzyme evolution by random recombinantion of improved sequences. *J. Mol. Biol.* 272, 336-347.

Murphy, P. J., Trenz, S. P., Grzemski, W., De Bruijn, F. J. and Schell, J. (1993) The *Rhizobium meliloti* rhizopine *mos* locus is a mosaic structure facilitating its symbiotic regulation. *J. Bacteriol.* 175, 5193-5204.

Nicolaus, B., de Simone, A., del Piano, L., Giardina, P. and Lama, L. (1986) Production of 2-keto-3-deoxygluconate by immobilised cells of *Sulfolobus solfataricus. Biotechnol. Lett.* 8, 497-500.

Nigam, P. and Singh, D. (1995) Enzyme and microbial systems involved in starch processing. *Enzyme Microb. Technol.* 17, 770-778.

O'Connell, E. L. and Meloche, H. P. (1982) Enzymic synthesis of 2-keto-3-deoxygluconate 6-phosphate using 6-phosphogluconate dehydratase. *Meth. Enzymol.* 89, 98-101.

Ohta, Y., Watanbe, K. and Kimura, A. (1985) Complete nucleotide sequence of the *E. coli* N-acetylneuraminate lyase. *Nucleic Acid Res.* 13, 8843-8852.

Okuta, A., Ohnishi, K. and Harayama, S. (1998) PCR isolation of catechol 2,3-dioxygenase gene fragments from environmental samples and their assembly into functional genes. *Gene.* 212, 221-228.

Otwinowski, Z, and Minor, W. (1996) Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* 276, 307-326.

Pace, C. N. (1992) Contribution of the hydrophobic effect to globular protein stability. *J. Mol. Biol.* 226, 29-35.

Pace, N. R. (1997) A molecular view of microbial diversity and the biosphere. *Science.* 276, 734-740.

Perutz, M. F. and Raidt, H. (1975) Stereochemical basis of heat stability in bacterial ferrodoxins and in haemoglobin A2. *Nature.* 255, 256-259.

Pieters, B. R. and Bardeletti, G. (1992) Enzyme immobilization on a low-cost magnetic support: kinetic studies on immobilized and coimmobilized glucose oxidase and glucoamylase. *Enzyme Microb. Technol.* 14, 361-370.

Plantier-Royon, R., Anker, D. and Robert-Baudouy, J. (1991) Nouvelle synthese de l'acide 3-desoxy-D-*erythro*-2-hexulosonique (KDG) à partir du D-glucose. *J. Carbohydr. Chem.* 10, 239-249.

Pollard, J. R., Rialland, D. and Bugg, T. D. H. (1998) Substrate selectivity and biochemical properties of 4-hydroxy-2-keto-pentanoic acid aldolase from *Escherichia coli*. *Appl. Environ. Microbiol.* 64, 4093-4094.

Portsmouth, D. (1968) Synthesis and properties of 3,6-dideoxyhexulosonic acids and related compounds. *Carbohyd. Res.* 8, 193-204.

Price, N. C. and Stevens, L. (1989) Fundamentals of Enzymology. Oxford University Press Inc., New York. 2nd edition. pp. 158-159.

Promega (1996) Protocols and Applications Guide. 3rd edition.

Qiao, L., Murray, B., Shimazaki, M., Schultz, J. and Wong, C-H. (1996) Synergistic inhibition of human a-1,3-fucosyltransferase V. *J. Am. Chem. Soc.* 118, 7653-7662.

Ramage, R., MacLeod, A. M. and Rose, G. W. (1991) Dioxalanones as synthetic intermediates. Part 6. Synthesis of 3-deoxy-D-*manno*-2-octulosonic acid (KDO), 3-deoxy-D-*arabino*-2-heptulosonic acid (DAH) and 2-keto-3-deoxy-D-gluconic acid (KDG). *Tetrahedron.* 47, 5625-5636.

Rao, J. P., Grzemski, W. and Murphy, P. J. (1995) *Rhizobium meliloti* lacking *mosA* synthesizes the rhizopine *scyllo*-inosamine in place of 3-*o*-methyl-*scyllo*-inosamine. *Microbiology.* 141:1683–1690.

Rhodes, G. (1993) In *Crystallography made crystal clear. A guide for users of macromolecular structures.* Academic Press, Inc., London.

Rosenberg, A. H., Lade, B. N., Chui, D-S, Lin, S-W., Dunn, J. J. and Studier, F. W. (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene.* 56, 125-135.

Russell, R. J. M., Ferguson, J. M. C., Hough, D. W., Danson, M. J. and Taylor, G. L. (1997) The crystal structure of citrate synthase from the hyperthermophilic archaeon *Pyrococcus furiosus* at 1.9 Å resolution. *Biochemistry.* 36, 9983-9994.

Russell, R. J. M., Gerike, U., Danson, M. J., Hough, D. W. and Taylor, G. L. (1998) Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. *Structure.* 6, 351-361.

Russell, R. J. M., Hough, D. W., Danson, M. J. and Taylor, G. L. (1994) The crystal structure of citrate synthase from the thermophilic Archaeon, *Thermoplasma acidophilum.* 2, 1157-1167.

Rutter, W. J. (1964) Evolution of aldolase. *Fed. Proc.* 23, 1248-1257.

Scopes, R. K. (1984) Use of differential dye-ligand chromatography with affinity elution for enzyme purification: 2-keto-3-deoxy-6-phosphogluconate aldolase from *Zymomonas mobilis*. *Anal. Biochem.* 136, 525-529.

Schauer, R. (1982) Chemistry, metabolism and biological functions of sialic acids. *Adv. Carbohydr. Chem. Biochem.* 40, 131-234.

Segerer, A. H. and Stetter, K. O. (1991) The order Sulfolobales. In *The Procaryotes*. Springer Verlag. (Balous, A et al., Ed.) 2nd edition. Vol. 1.

Selig, M., Xavier, K. B., Santos, H. and Schönheit, P. (1997) Comparative analysis of Embden-Meyerhof and Entner-Doudoroff glycolytic pathways in hyperthermophilic archaea and the bacterium *Thermotoga*. *Arch. Microbiol.* 167, 217-232.

Sellek, G. A. and Chaudhuri. J. B. (1999) Biocatalysis in organic media using enzymes from extremophiles. *Enzyme Microb. Technol.* 25, 471-482.

Shedlarski, J. G. and Gilvarg, C. (1970) The pyruvate-aspartic semialdehyde condensing enzyme of *Escherichia coli*. *J. Biol. Chem.* 245, 1362-1373.

Shelton, M. C., Cotterill, I. C., Novak, S. T. A., Poonawala, R. M., Sudarshan, S. and Toone, E. J. (1996) 2-Keto-3-deoxy-6-phosphogluconate aldolases as catalysts for stereocontrolled carbon-carbon bond formation. *J. Am. Chem. Soc.* 118, 2117-2125.

Shimizu, M., Yoshida, A. and Mikami, K. (1996) Carbonyl-ene approach to the assymetric synthesis of 2-keto-3-deoxy-D-gluconic acid (KDG): a combinatorial sequence using Sharpless epoxidation. *Synlett.* 11, 1112-1114.

Shirakawa, M., Tsurimoto, T. and Matsubara, K. (1984) Plasmid vectors designed for high efficiency expression controlled by the portable *recA* promoter-operator of *Escherichia coli*. *Gene* 28, 127-132.

Simpson, F. J., Perlin, A. S. and Sieben, A. S. (1966) Erythrose 4-phosphate. *Meth. Enzymol.* 9, 35-38.

Skoza, L. and Mohos, S. (1976) Stable thiobarbituric acid chromophore with dimethyl sulphoxide. Application to sialic acid assay ion analytical de-*O*-acetylation. *Biochem. J.* 159, 457-462.

Sparks, M. A., Williams, K. W., Lukacs, C., Schrell, A., Priebe, G., Spaltenstein, A. and Whitesides, G. M. (1993) Synthesis of potential inhibitors of haemagglutinin by influenza virus: chemoenzymic preparation of N-5 analogs of *N*-acetylneuraminic acid. *Tetrahedron.* 49, 1-12.

Starr, M. P. and Chatterjee, A. K. (1972) The genus *Erwinia*: enterobacteria pathogenic to plants and animals. *Annu. Rev. Microbiol.* 26, 389-426.

Studier, F. W. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* 219, 37-44.

Studier, F. W. and Moffatt, B. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* 189, 113-130.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. and Dubendorff, J. W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Meth. Enzymol.* 185, 60-89.

Sugai, T., Shen, G-J., Ichikawa, Y. and Wong, C-H. (1993) Synthesis of 3-deoxy-D-*manno*-2-octulosonic acid (KDO) and its analogs based on KDO aldolase-catalysed reactions. *J. Am. Chem. Soc.* 115, 413-421.

Takayama, S., McGarvey, G. J. and Wong, C-H. (1997) Microbial aldolases and transketolases: new biocatalytic approaches to simple and complex sugars. *Annu. Rev. Microbiol.* 51, 285-310.

Tanner., J. J. Hecht, R. M. and Krause, K. L. (1996) Determinants of enzyme thermostability observed in the molecular structure of *Thermus aqauticus* D-glyceraldehyde-3-phosphate dehydrogenase at 2.5 Å resolution. *Biochemistry.* 35, 2597-2609.

Thomas, T. M. and Scopes, R. K. (1998) The effects of temperature on the kinetics and stability of mesophilic and thermophilic 3-phosphoglycerate kinases. *Biochem. J.* 330, 1087-1095.

Turner, M. M. (1995) Biocatalysis in organic chemistry (Part II): present and future. *Trends Biotechnol.* 13, 253-258.

Uchida, Y., Tsukada, Y. and Sugimori, T. (1984) Purification and properties of *N*-acetylneuraminate lyase from *Escherichia coli*. *J. Biochem.* 96, 507-522.

Vieille, C., Hess, J. M., Kelly, R. M. and Zeikus, J. G. (1995) *xylA* Cloning and sequencing and biochemical characterisation of xylose isomerase from *Therotoga neapolitana*. *Appl. Environ. Microbiol.* 61, 1867-1875.

Vimr, E. R. and Troy, F. A. (1985a) Identification of an inducible catabolic system for sialic acids (*nan*) in *Escherichia coli*. *J. Bacteriol.* 164, 845-853.

Vimr, E. R. and Troy, F. A. (1985b) Regulation of sialic acid metabolism in *Escherichia coli*: role of *N*-acylneuraminate pyruvate-lyase. *J. Bacteriol.* 164, 854-860.

Weissback, A. and Hurwitz, J. (1959) The formation of 2-keto-3-deoxyheptonic acid in extracts of *Escherichia coli* B. *J. Biol. Chem.* 234, 705-709.

West, S. I. (1988) Enzymes in the food processing industry. *Chemistry in Britain.* 24, 1220-1222.

Wharton, C. W. and Eisenthal, R. (1981) Molecular Enzymology. Blackie and Son, Ltd., Glasgow and London, UK. p. 141.

Witke, C. and Götz, F. (1993) Cloning, sequencing, and characterisation of the gene encoding the class I fructose-1,6-bisphosphate aldolase of *Staphylococcus carnosus*. *J. Bacteriol.* 175, 7495-7499.

Woese, C. R. and Fox, G. E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA.* 74, 5088-5090.

Woese, C. R., Kandler, O. and Wheelis, M. L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya. *Proc. Natl. Acad. Sci. USA.* 87, 4576-4579.

Woese, C. R. and Olsen, G. J. (1986) Archaebacterial phylogeny: perspectives on the Urkingdoms. *System. Appl. Microbiol.* 7, 161-177.

Yeats, S., McWilliam, P. and Zillig, W. (1982) A plasmid in the archaebacterium *Sulfolobus acidocaldarius*. *Embo J.* 1, 1035-1038.

You, L. and Arnold, F. H. (1994) Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Eng.* 9, 77-83.

APPENDIX

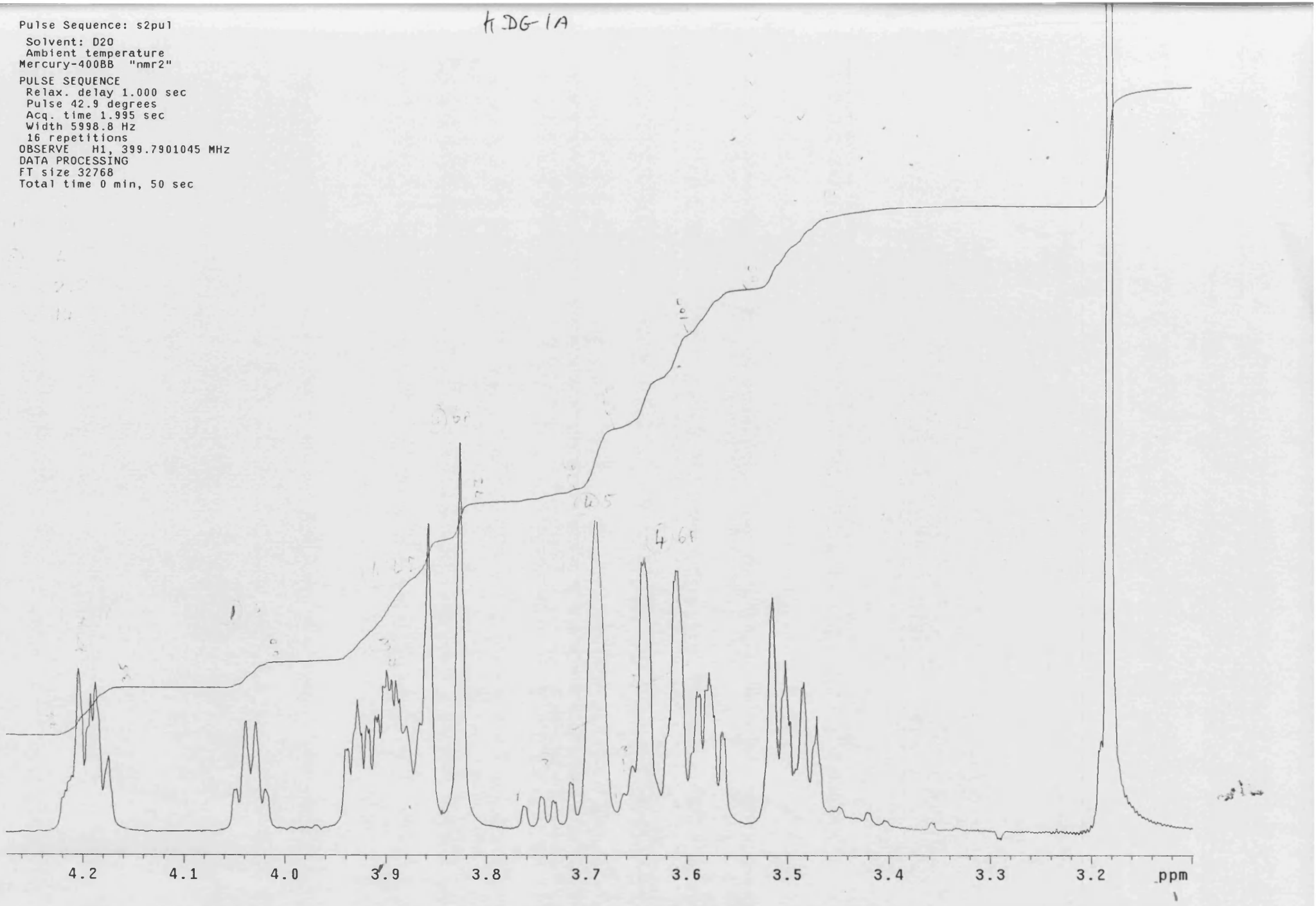
I

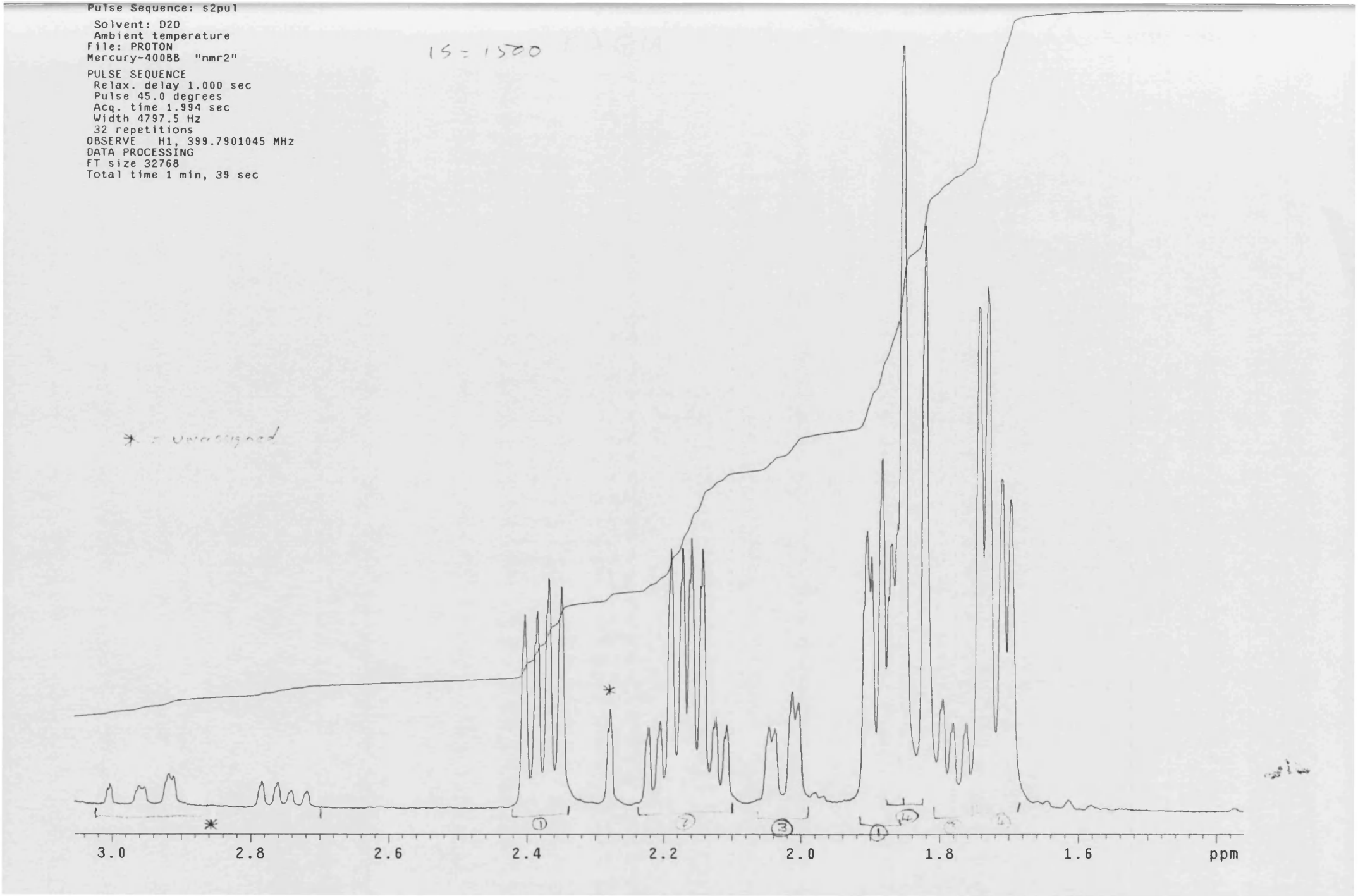
^{1}H NMR SPECTRA FOR KDG POOL 1

HDG-1A

Pulse Sequence: s2pul
Solvent: D2O
Ambient temperature
Mercury-400BB "nmr2"

PULSE SEQUENCE
Relax. delay 1.000 sec
Pulse 42.9 degrees
Acq. time 1.995 sec
Width 5998.8 Hz
16 repetitions
OBSERVE H1, 399.7901045 MHz
DATA PROCESSING
FT size 32768
Total time 0 min, 50 sec

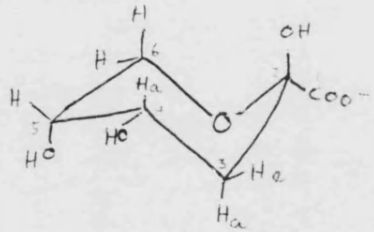



Pulse Sequence: s2pul
Solvent: D2O
Ambient temperature
File: PROTON
Mercury-400BB "nmr2"
PULSE SEQUENCE
Relax. delay 1.000 sec
Pulse 45.0 degrees
Acq. time 1.994 sec
Width 4797.5 Hz
32 repetitions
OBSERVE H1, 399.7901045 MHz
DATA PROCESSING
FT size 32768
Total time 1 min, 39 sec
IS = 1500
* = Unassigned
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
ppm

Pulse Sequence: s2pul
Solvent: D2O
Ambient temperature
Mercury-400BB "nmr2"
PULSE SEQUENCE
Relax. delay 1.000 sec
Pulse 42.9 degrees
Acq. time 1.995 sec
Width 5998.8 Hz
16 repetitions
OBSERVE H1, 399.7901045 MHz
DATA PROCESSING
FT size 32768
Total time 0 min, 50 sec

KDG1A

(4)



β - pyr

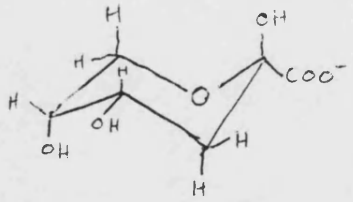
STANDARD 1H OBSERVE

Pulse Sequence: s2pul
Solvent: D2O
Ambient temperature
Mercury-400BB "nmr2"

PULSE SEQUENCE
Relax. delay 1.000 sec
Pulse 42.9 degrees
Acq. time 1.995 sec
Width 5998.8 Hz
16 repetitions
OBSERVE H1, 399.7901045 MHz
DATA PROCESSING
FT size 32768
Total time 0 min, 50 sec

HDG-1A

(4)



β - pyr

6A (13)

6B

5

4

4.2 4.1 4.0 3.9 3.8 3.7 3.6 3.5 3.4 3.3 3.2 ppm

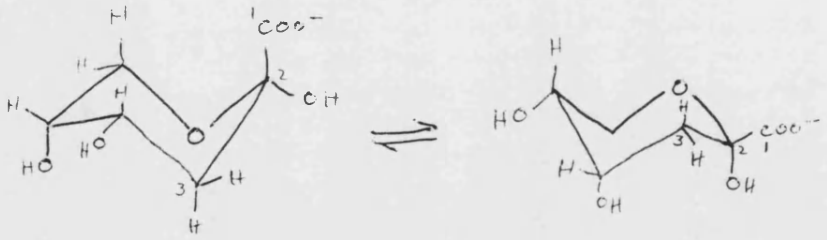
STANDARD 1H OBSERVE

Pulse Sequence: s2pul
Solvent: D2O
Ambient temperature
Mercury-400BB "nmr2"

PULSE SEQUENCE
Relax. delay 1.000 sec
Pulse 42.9 degrees
Acq. time 1.995 sec
Width 5998.8 Hz
16 repetitions
OBSERVE H1, 399.7901045 MHz
DATA PROCESSING
FT size 32768
Total time 0 min, 50 sec

HTDG1A

(3)



α - pyr

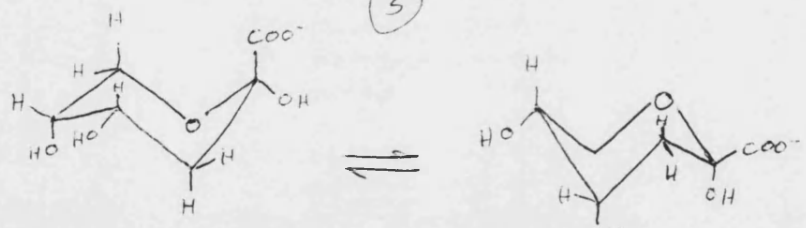
STANDARD 1H OBSERVE

Pulse Sequence: s2pul
Solvent: D2O
Ambient temperature
Mercury-400BB "nmr2"

PULSE SEQUENCE
Relax. delay 1.000 sec
Pulse 42.9 degrees
Acq. time 1.995 sec
Width 5998.8 Hz
16 repetitions
OBSERVE H1, 399.7901045 MHz
DATA PROCESSING
FT size 32768
Total time 0 min, 50 sec

HDG-1A

(3)



α-pyr

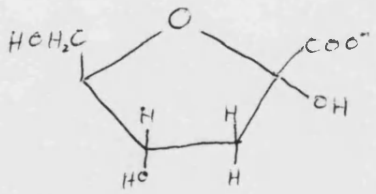
STANDARD 1H OBSERVE

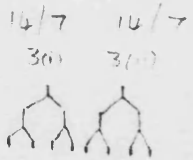
Pulse Sequence: s2pul
Solvent: D2O
Ambient temperature
Mercury-400BB "nmr2"

PULSE SEQUENCE
Relax. delay 1.000 sec
Pulse 42.9 degrees
Acq. time 1.995 sec
Width 5998.8 Hz
16 repetitions
OBSERVE H1, 399.7901045 MHz
DATA PROCESSING
FT size 32768
Total time 0 min, 50 sec

KDG1A

(2)





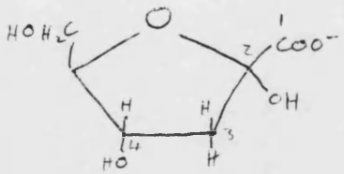
Pulse Sequence: s2pul

Solvent: D2O
Ambient temperature
Mercury-400BB "nmr2"

PULSE SEQUENCE
Relax. delay 1.000 sec
Pulse 42.9 degrees
Acq. time 1.995 sec
Width 5998.8 Hz
16 repetitions
OBSERVE H1, 399.7901045 MHz
DATA PROCESSING
FT size 32768
Total time 0 min, 50 sec

ADG-1A

(2)



4

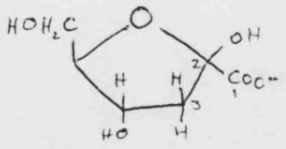
STANDARD 1H OBSERVE

Pulse Sequence: s2pul
Solvent: D2O
Ambient temperature
Mercury-400BB "nmr2"

PULSE SEQUENCE
Relax. delay 1.000 sec
Pulse 42.9 degrees
Acq. time 1.995 sec
Width 5998.8 Hz
16 repetitions
OBSERVE H1, 399.7901045 MHz
DATA PROCESSING
FT size 32768
Total time 0 min, 50 sec

KDG1A

(1)



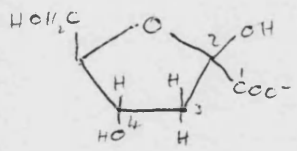
STANDARD 1H OBSERVE

Pulse Sequence: s2pul
Solvent: D2O
Ambient temperature
Mercury-400BB "nmr2"

PULSE SEQUENCE
Relax. delay 1.000 sec
Pulse 42.9 degrees
Acq. time 1.995 sec
Width 5998.8 Hz
16 repetitions
OBSERVE H1, 399.7901045 MHz
DATA PROCESSING
FT size 32768
Total time 0 min, 50 sec

HDG-1A

(1)



4

4.2 4.1 4.0 3.9 3.8 3.7 3.6 3.5 3.4 3.3 3.2 ppm

APPENDIX

II

PUBLICATION

Buchanan, C. L.,[1] Connaris, H., Danson, M. J., Reeve, C. D. and Hough, D. W. (1999) An extremely thermostable aldolase from *Sulfolobus solfataricus* with specificity for non-phosphorylated substrates. *Biochem. J.* 342, 563-570.

[1] The author Catriona L. Buchanan is also known as Catriona L. Kydd.

Biochem. J. (1999) 343, 563–570 (Printed in Great Britain)

An extremely thermostable aldolase from *Sulfolobus solfataricus* with specificity for non-phosphorylated substrates

Catriona L. BUCHANAN*, Helen CONNARIS*, Michael J. DANSON*, Christopher D. REEVE† and David W. HOUGH*[1]

*Centre for Extremophile Research, Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, U.K., and †Avecia LifeScience Molecules, Belasis Avenue, Billingham, Cleveland TS23 1YN, U.K.

Sulfolobus solfataricus is a hyperthermophilic archaeon growing optimally at 80–85 °C. It metabolizes glucose via a novel non-phosphorylated Entner–Doudoroff pathway, in which the reversible C_6 to C_3 aldol cleavage is catalysed by 2-keto-3-deoxygluconate aldolase (KDG-aldolase), generating pyruvate and glyceraldehyde. Given the ability of such a hyperstable enzyme to catalyse carbon–carbon-bond synthesis with non-phosphorylated metabolites, we report here the cloning and sequencing of the *S. solfataricus* gene encoding KDG-aldolase, and its expression in *Escherichia coli* to give fully active enzyme. The recombinant enzyme was purified in a simple two-step procedure, and shown to possess kinetic properties indistinguishable from the enzyme purified from *S. solfataricus* cells. The KDG-aldolase is a thermostable tetrameric protein with a half-life at 100 °C of 2.5 h, and is equally active with both D- and L-glyceraldehyde. It exhibits sequence similarity to the *N*-acetylneuraminate lyase superfamily of Schiff-base-dependent aldolases, dehydratases and decarboxylases, and evidence is presented for a similar catalytic mechanism for the archaeal enzyme by substrate-dependent inactivation by reduction with $NaBH_4$.

Key words: biotransformation, C–C bond, 2-keto-3-deoxygluconate.

INTRODUCTION

Sulfolobus solfataricus is a hyperthermophilic Archaeon that grows optimally at 80–85 °C and pH 2–4 [1]. With respect to its pathways of central metabolism, glucose is oxidized via a non-phosphorylated Entner–Doudoroff pathway (Scheme 1) [2], involving an NAD(P)-dependent dehydrogenation of glucose to gluconate that is then dehydrated to form 2-keto-3-deoxygluconate (KDG). KDG then undergoes an aldol cleavage catalysed by KDG-aldolase (EC 4.1.2.20):

2-Keto-3-deoxygluconate ⇌ pyruvate + glyceraldehyde

Up to this point, there is no requirement for ATP. After the conversion of glyceraldehyde to glycerate, the action of a specific kinase produces 2-phosphoglycerate, which is further converted to generate a second molecule of pyruvate via enolase and pyruvate kinase.

This non-phosphorylated Entner–Doudoroff pathway has also been reported in the Archaeon, *Thermoplasma acidophilum* [3], and is used for glucose catabolism to a small extent in *Thermoproteus tenax* [4], but not in *Pyrococcus furiosus*, even though this organism contains the majority of the necessary enzymes (reviewed in [5]). This situation contrasts with that in bacteria, where the Entner–Doudoroff pathway, if present, involves phosphorylated metabolites.

In the current article, we report the purification of the *S. solfataricus* KDG-aldolase, which then permitted the cloning and expression of the gene in *Escherichia coli*, followed by characterization of the recombinant enzyme. The potential use of KDG-aldolase in its synthetic mode may overcome one of the most challenging tasks in synthetic chemistry, namely the formation of the carbon–carbon bond. Furthermore, its inherent thermostability and ability to utilize non-phosphorylated substrates may offer significant advantages over currently available enzyme systems.

EXPERIMENTAL

Strains and plasmids

Cells of *S. solfataricus* (DSM 1616) were kindly provided by Dr. Neil Raven (Centre for Applied Microbiological Research, Porton Down, Wilts., U.K.). PCR products were cloned into the pGEM-T vector system (Promega, Madison, WI, U.S.A.). The expression vector pREC7/*Nde*I (provided kindly by Dr. L. C. Kurz, Washington University School of Medicine, St. Louis, MO, U.S.A.) was used for expression in the *E. coli* strain JM109 (Promega).

Purification of KDG-aldolase from *S. solfataricus*

Preparation of *S. solfataricus* cell extracts

Cell paste was resuspended in 20 mM Tris/HCl, pH 8.5, containing 1 mM PMSF and 1 mM EDTA, at approx. 0.2 g of cells/ml. Cells were broken by four 30-s bursts of sonication on ice with an MSE 150-W Ultrasonic Disintegrator. Soluble cell extract was obtained by centrifugation at 25000 *g* for 1 h at 4 °C. The insoluble pellet was discarded.

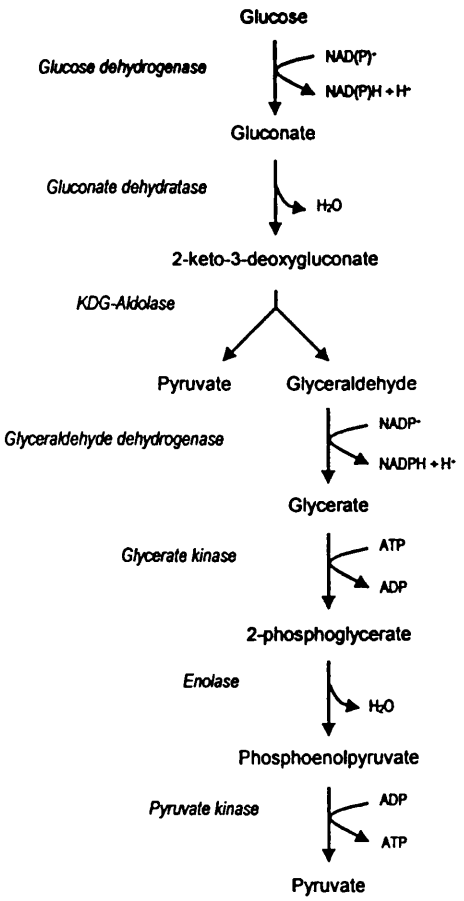
Heat step

The soluble cell extract was heated in glass test tubes (1-cm diameter) in a water bath at 90 °C for 15 min. Precipitated proteins were then removed by centrifugation at 13000 *g*, for 1 h, at 4 °C.

Abbreviations used: TBA, thiobarbituric acid; KDG, 2-keto-3-deoxygluconate (2-oxo-3-deoxygluconate); DHDPS, dihydrodipicolinate synthase; NAL, *N*-acetylneuraminate lyase; ORF, open reading frame.

[1] To whom correspondence should be addressed (e-mail D.W.Hough@bath.ac.uk).

The nucleotide sequence data reported have been deposited in the GenBank sequence data base under the accession number AJ224174.



Scheme 1 The non-phosphorylated Entner–Doudoroff pathway of *S. solfataricus*

Ammonium sulphate precipitation

Pulverized ammonium sulphate was added to the heat-treated cell extract at 4 °C to a concentration of 65% (w/v), and the solution was maintained at this temperature for 15 min. Precipitated proteins were then removed by centrifugation at 11000 *g* for 20 min at 4 °C, and the supernatant containing the KDG-aldolase activity was retained and further purified by gel filtration.

Gel filtration and ion-exchange chromatography

Gel filtration was carried out on a Sephacryl S-300 HR column (5 cm × 44 cm) in 20 mM Tris/HCl buffer, pH 8.5. A HiTrap Q anion-exchange column was equilibrated in 20 mM Tris/HCl, pH 8.5. After application of the gel-filtration fractions containing KDG-aldolase activity, the column was washed with the same buffer and the bound protein was then eluted using a gradient of 0–2 M NaCl in 20 mM Tris/HCl, pH 8.5.

Preparative PAGE

Partially purified enzyme following anion-exchange chromatography was electrophoresed in three adjacent lanes of a non-denaturing polyacrylamide gel. Gel buffers were prepared as for the subsequent SDS/PAGE, but with the omission of SDS. One lane was stained with Coomassie Brilliant Blue. The remaining two lanes were cut horizontally into 1-mm fractions; slices from one of these lanes were ground in 20 μl of SDS/PAGE loading dye and then subjected to SDS/PAGE [6]. Slices from the second lane were ground in 50 mM sodium phosphate buffer, pH 6.0, and the resulting extract was assayed for KDG-aldolase activity using the thiobarbituric acid (TBA) assay as described below.

Analytical techniques

TBA assay of KDG-aldolase activity

KDG-aldolase activity was measured using a modification of the TBA assay [7,8]. Reactions of total volume 250 μl were incubated at 70 °C in 50 mM sodium phosphate buffer, pH 6.0, with 50 mM pyruvate and 20 mM D,L-glyceraldehyde and an appropriate volume of cell extract, or purified enzyme, between 10 and 50 μl. After 10 min, 100-μl samples were removed and the reaction stopped by the addition of 10 μl of 12% (w/v) trichloroacetic acid. Precipitated proteins were removed by centrifugation, and 50 μl of the supernatant (containing the product of the KDG-aldolase, KDG) were then oxidized by the addition of 125 μl of 25 mM periodic acid/0.25 M H_2SO_4 and incubation at room temperature for 20 min. Oxidation was terminated by the addition of 250 μl of 2% (w/v) sodium arsenite in 0.5 M HCl. TBA [1 ml, 0.3% (w/v)] was then added and the chromophore developed by heating at 100 °C for 10 min. A sample of this solution was then removed and the colour intensified by adding to an equal volume of DMSO. The absorbance was read at 549 nm, where the absorbance coefficient for the chromophore was taken to be 67.8×10^3 $M^{-1} \cdot cm^{-1}$ [7,8]. All enzyme activities were determined in duplicate, and the S.E.M. of reported activity values was determined to be $\pm 3.6\%$.

HPLC assay of KDG-aldolase activity

Samples from aldolase assay reactions (as above) were analysed by HPLC using a Bio-Rad Aminex HPX-87H Organic Analysis column (300 mm × 7.8 mm) linked to a refractive-index detector. Reaction samples (5 μl) were injected and eluted with 8 mM H_2SO_4 at a flow rate of 1 ml/min. The system was calibrated with pyruvate, glyceraldehyde and KDG.

Protein-concentration determination

Protein concentrations were determined by the method of Bradford [9] using a calibration curve constructed with BSA.

Protein microsequencing

Following preparative gel electrophoresis, the band corresponding to KDG-aldolase activity was electroblotted on to a hydrophobic PVDF membrane. The N-terminal sequence was then determined on an Applied Biosystems 470 gas-phase sequencer, coupled to an Applied Biosystems 120 phenylthiohydantoin analyser.

Gene cloning and sequencing

Degenerate PCR primers were designed based on each end of the 38-residue N-terminal amino acid sequence obtained from the purified KDG-aldolase. PCR, using *S. solfataricus* genomic DNA as template, yielded a 114-bp product that was then radiolabelled using a High Prime DNA labelling kit (Boehringer-Mannheim, Lewes, Sussex, U.K.). This was then used as a hybridization probe to screen a *S. solfataricus* genomic DNA library in lambda phage [10]. Screening was carried out as described by Sambrook et al. [11], and a secondary screen was performed on initial positives. Lambda DNA from secondary positives was purified

using the Qiagen lambda kit, and DNA sequencing was carried out using the dideoxynucleotide chain-termination method [12] on an Applied Biosystems 377 automated DNA sequencer.

Expression and purification of recombinant enzyme

Expression

The aldolase gene from the lambda library was amplified by PCR using a 5′ primer that introduced a unique *Nde*I site upstream of the initiating methionine codon and a 3′ primer that introduced a unique *Bam*HI site downstream of the termination codon. The PCR product was digested with the appropriate restriction endonucleases and cloned into *Nde*I/*Bam*HI-digested pREC7 expression vector; this was then used to transform *E. coli* strain JM109. Luria broth medium (1 litre) containing carbenicillin (50 μg/ml) was inoculated with 50 ml of an overnight culture of transformed JM109 cells, and the culture was then grown at 37 °C to a D_{600} of 0.6. The cells were then induced with nalidixic acid (50 μg/ml) and grown for a further 21 h at 37 °C. Cells were harvested at 2000 *g* for 10 min and the cell pellet stored at −20 °C.

Purification of recombinant KDG-aldolase

Cell paste was re-suspended in 20 mM Tris/HCl, pH 8.5, containing 1 mM PMSF and 1 mM EDTA, at approx. 0.2 g of cells/ml. Triton X-100 (final concentration 0.1%, w/v) and lysozyme (100 μg/ml) were added, and the cell suspension incubated at 37 °C with occasional shaking for 1 h. Cell lysis was observed during this time. Complete lysis was achieved by five 30-s bursts of sonication on ice with an MSE 150-W Ultrasonic Disintegrator, after which the cell debris was removed by centrifugation at 11000 *g* for 30 min at 4 °C.

The cell extract was heated in glass test tubes (1-cm diameter) in a water bath at 78 °C for 30 min, after which precipitated proteins were removed by centrifugation at 11000 *g* for 20 min at 4 °C. The supernatant was then applied to a 2.6 cm × 1.5 cm column of Q Sepharose (Pharmacia) equilibrated with 20 mM Tris/HCl, pH 8.5. The column was washed with equilibration buffer to remove unbound protein, and then bound protein was eluted with a gradient of 0–2 M NaCl in 20 mM Tris/HCl, pH 8.5. Fractions (2 ml) were collected and analysed for catalytic activity and for purity by SDS/PAGE.

Characterization of the recombinant KDG-aldolase

Thermostability

Enzyme assay buffer, overlaid with mineral oil to minimize evaporation, was incubated at 100 °C in 0.5-ml sealed tubes. Once the correct temperature was reached, one-fiftieth volume of enzyme was added to each tube. At selected times, a tube was removed and frozen by plunging into a solid-CO_2/methanol bath. Tubes were stored at −20 °C until assayed for KDG-aldolase activity using the TBA assay described above.

Determination of relative molecular mass (M_r)

The M_r value of the KDG-aldolase was determined by centrifugal analysis on a Beckman XL-A analytical ultracentrifuge at the U.K. National Centre for Macromolecular Hydrodynamics, University of Nottingham, Nottingham, U.K. Sedimentation equilibrium experiments were carried out at 20 °C and three rotor speeds (10000, 15000 and 20000 revs./min) were employed. Three solute concentrations (0.1, 0.3 and 1 mg of protein/ml) were used, and data were captured at 236 and 280 nm, using the scanning absorption optical system. Data were analysed using the program NONLIN [13], fitting to both single species and associating system models. Sedimentation velocity experiments were carried out at 20 °C and 40000 rev./min. Three solute concentrations (0.4, 0.6 and 0.8 mg of protein/ml) were used, and data were analysed using the program Svedberg (Amgen, Thousand Oaks, CA, U.S.A.).

The diffusion coefficient of the KDG-aldolase was determined by dynamic light scattering at 20 °C on a DynaPro 801 Dynamic Light Scattering Instrument (Protein Solutions, High Wycombe, Bucks, U.K.), using a sample at 2 mg of protein/ml [14].

Reduction with $NaBH_4$

A method similar to that used by Aisaka et al. ([15], and references therein) was used to investigate the possible involvement of a Schiff-base mechanism in the KDG-aldolase reaction. Enzyme was incubated at room temperature in 50 mM sodium phosphate buffer, with 100 mM $NaBH_4$ in the presence or absence of saturating concentrations of pyruvate or D,L-glyceraldehyde (10 and 20 mM, respectively). After 10 min the samples were dialysed twice against 2 l of 20 mM Tris/HCl, pH 8.5, and assayed for KDG-aldolase activity in the TBA assay.

RESULTS

Purification of native KDG-aldolase

The purification schedule used to obtain KDG-aldolase for N-terminal sequencing is summarized in Table 1. Briefly, as described in the Experimental section, purification from cell extracts involved a heat treatment, an ammonium sulphate fractionation, and chromatography by gel filtration and anion exchange. Samples after each step were analysed for catalytic activity, and for purity by SDS/PAGE. Approx. 15 μg of protein from the peak fraction following anion-exchange chromatography, which had a specific activity of 1.9 units/mg, was subjected to preparative PAGE. When the gel was analysed as described in the Experimental section, a major band of $M_r \approx$ 32000 was identified as the only protein that possessed KDG-aldolase activity. Protein from this band was eluted, subjected to SDS/PAGE and blotted on to PVDF membrane. The major band was excised and its N-terminal sequence was determined to be: PEIIPIITPFTKDNRIDKEKLKIHAENLIRKGIDKLF.

Cloning of the KDG-aldolase gene from *S. solfataricus*

Amplification of *S. solfataricus* genomic DNA using oligonucleotide primers based on the 38-amino-acid N-terminal sequence yielded a 114-bp fragment. This was cloned into pGEM-T vector without further modification and sequenced. As expected, the sequence was found to contain an insert with a nucleotide sequence corresponding to the determined N-terminal amino acid sequence of the KDG-aldolase. The 114-bp PCR fragment was radiolabelled and used to screen an *S. solfataricus* genomic DNA library. A small number of the positive plaques obtained were used in a secondary screen. Lambda DNA from a positive plaque from the secondary screen was purified, and the sequence of the KDG-aldolase gene was obtained by direct sequencing of both strands.

Nucleotide sequence analysis

The complete nucleotide sequence and deduced amino acid sequence of the *S. solfataricus* KDG-aldolase gene are shown in

Table 1 Purification of KDG-aldolase from *S. solfataricus*

Step	Protein (mg)	Total activity (units)	Specific activity (units/mg)	Yield (%)	Fold purification
Cell homogenate	1420	34.7	0.025	100	1
Soluble cell extract	1220	32.5	0.027	93	1
90 °C heat step	1040	28.1	0.027	81	1
65% (w/v) ammonium sulphate supernatant	71	16.6	0.24	71	10
Sephacryl-300 gel filtration	7.4	4.1	0.55	12	23
HiTrap Q anion exchange	2.7	2.9	1.1	8	44

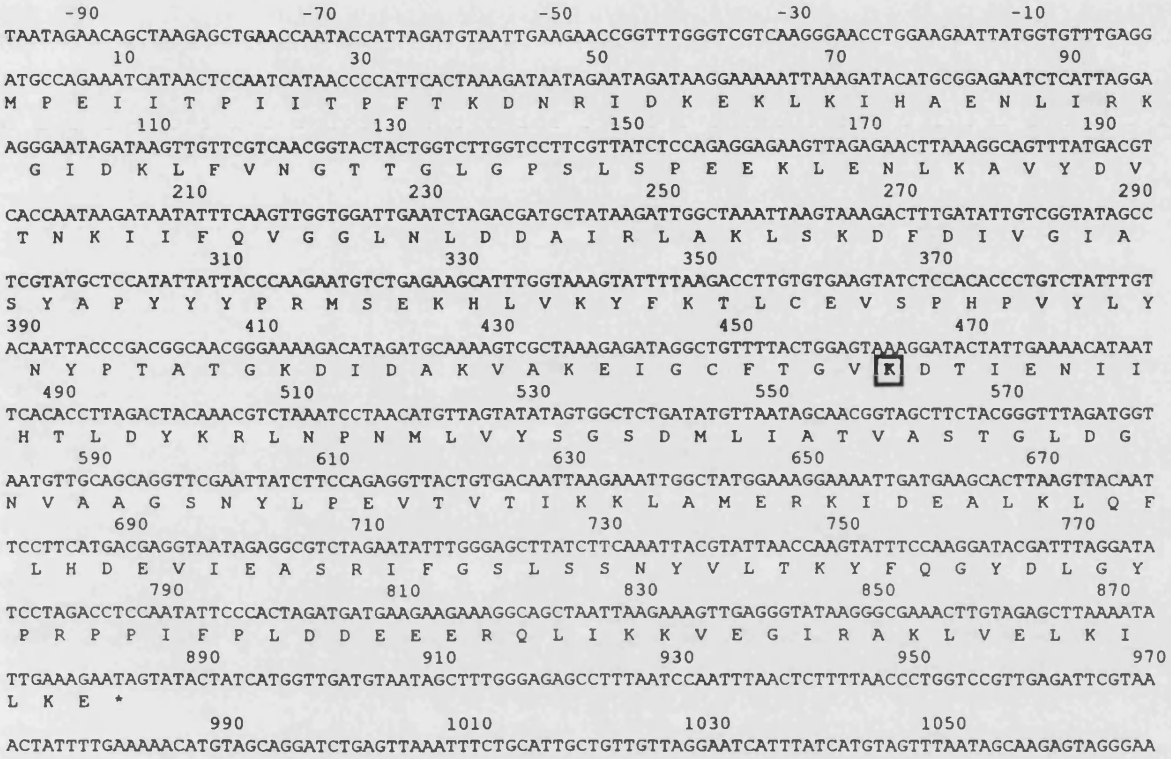


Figure 1 Nucleotide sequence of the KDG-aldolase gene from *S. solfataricus* and its flanking regions

The nucleotide sequence of a 1165-bp fragment is shown, with the region encoding the KDG-aldolase numbered from 1–882 (including the ATG start codon); the deduced amino acid sequence is shown below this sequence in the one-letter code. The proposed active-site lysine residue (K155) is shown boxed and in bold type. The upstream (97-bp) and downstream (186-bp) nucleotide flanking sequences of the gene are also shown.

Figure 1. The ATG initiation codon was identified by the exact match of the deduced amino acid sequence following the methionine with that determined from the purified protein. The gene encodes a polypeptide of 293 amino acids with an M_r of 32980.

Database searches and sequence alignments were performed using Basic Local Alignment Search Tool (BLAST) [16]. The highest scores were observed with dihydrodipicolinate synthase (DHDPS; EC 4.2.1.52) from *Haemophilus influenzae* (30% identity), *Bacillus subtilis* (29%) and *E. coli* (28%), and the *N*-acetylneuraminate lyase (NAL; EC 4.1.3.3) of *H. influenzae* (28%) and *E. coli* (26%). Significant identities were also seen with the *Rhizobium meliloti* MosA protein (27%) and the probable 5-dehydro-4-deoxyglucarate dehydratase from *B. subtilis* (26%). Throughout these enzymes, an active-site lysine is found that forms a Schiff-base intermediate with the pyruvate substrate [17,18]. Despite its low overall sequence identity with these proteins, the *S. solfataricus* KDG-aldolase also has this conserved lysine residue (K155; Figure 2). Further evidence that KDG-aldolase is a member of the NAL superfamily is provided by the fact that it possesses three of the other five residues identified by Lawrence et al. [19] as strictly conserved in all members of the superfamily. Two of these (G45 and Y137) correspond to residues involved in stabilizing the intermediate enzyme–substrate complex, and the other (G201) occurs at the N-terminus of a β-strand.

Expression and purification of the *S. solfataricus* KDG-aldolase gene in *E. coli*

Recombinant KDG-aldolase was expressed using pREC7-transformed JM109 cells grown to logarithmic phase in Luria broth and induced with nalidixic acid for 21 h at 37 °C. The specific

			*
Aldolase	*S. solfataricus*	(148)	IGCFTGVKDTIENIIHT
NAL	*E. coli*	(158)	LPGVGALKQTSGDLYQM
	H. influenzae	(157)	NPKVLGVKFTAGDFYLL
DHDPS	*E. coli*	(154)	VKNIIGIKEATGNLTRV
	H. influenzae	(160)	IENIVGIKEATRDVSRI
	B. subtilis	(156)	IPNVVAIKEASGDLEAI

Figure 2 Comparison of the amino acid sequence of *S. solfataricus* KDG-aldolase with members of the NAL superfamily

NAL from *E. coli* (P06995) and *H. influenzae* (P44539), and DHDPS from *E. coli* (P05640), *H. influenzae* (P43797) and *B. subtilis* (Q04796) were aligned with *S. solfataricus* KDG-aldolase. Numbers in parentheses refer to the relevant entry in the SWISSPROT database; the alignment was produced using CLUSTAL [32]. The conserved active-site lysine residue is marked with an asterisk.

Table 2 Kinetic parameters of KDG-aldolase

Data were analysed using the direct linear plot [33]; values in parentheses are standard errors.

	K_m (mM)	V_{max} (units/mg)
Wild-type KDG-aldolase (semi-purified)		
D,L-Glyceraldehyde	3.6 (±0.2)	8.4 (±0.2)
Pyruvate	0.9 (±0.2)	6.3 (±0.2)
Recombinant KDG-aldolase (purified)		
D,L-Glyceraldehyde	5.2 (±0.1)	17.1 (±0.4)
D-Glyceraldehyde	3.9 (±0.3)	18.0 (±1.0)
L-Glyceraldehyde	7.1 (±0.6)	18.0 (±1.9)
Pyruvate	1.0 (±0.1)	15.7 (±0.3)

activity of the enzyme was 0.5 unit/mg of protein, a value 18-fold higher than endogenous activity in *S. solfataricus* cells. A heat step of 78 °C for 30 min followed by anion-exchange chromatography yielded enzyme with a specific activity of 9 units/mg of protein, which was shown to be homogeneous by SDS/PAGE.

Characterization of recombinant KDG-aldolase

Thermoactivity and thermostability

KDG-aldolase activity was measured in the direction of KDG synthesis using a modification of the TBA assay. The standard assay temperature in all studies reported in this paper was 70 °C, at which temperature the linearity of the assay with time was demonstrated for both native and recombinant enzymes, as was the measurement of initial rates. Using this assay, the recombinant protein was found to possess similar catalytic properties to the native enzyme, both proteins displaying Michaelis–Menten kinetics (Table 2). No activity could be detected with glyceraldehyde 3-phosphate as substrate in place of glyceraldehyde or with the pyruvate analogues β-hydroxypyruvate and α-ketobutyrate. It has been reported that β-hydroxypyruvate is an inhibitor of *E. coli* NAL [19], and we have found that this compound is also an inhibitor of *S. solfataricus* KDG-aldolase (K_i = 0.3 mM).

Using the purified recombinant enzyme, the KDG-aldolase was shown to have equal catalytic activity towards D- and L-glyceraldehyde, although there are minor differences in their respective K_m values. For assays carried out on the racemic mixture, D-glyceraldehyde will act as a competitive inhibitor of the L-glyceraldehyde, with a K_i equal to its K_m, and vice versa. In this situation, the rate equation describing the velocity (v) observed is:

$$v = \frac{(V^D_{max} \cdot S^D/K^D_m) + (V^L_{max} \cdot S^L/K^L_m}{1 + S^D/K^D_m + S^L/K^L_m}$$

where the concentrations (S) and kinetic constants (K_m and V_{max}) of the D- and L-glyceraldehydes are indicated by their respective superscripts ([20], p. 141). Given identical V_{max} values for the two

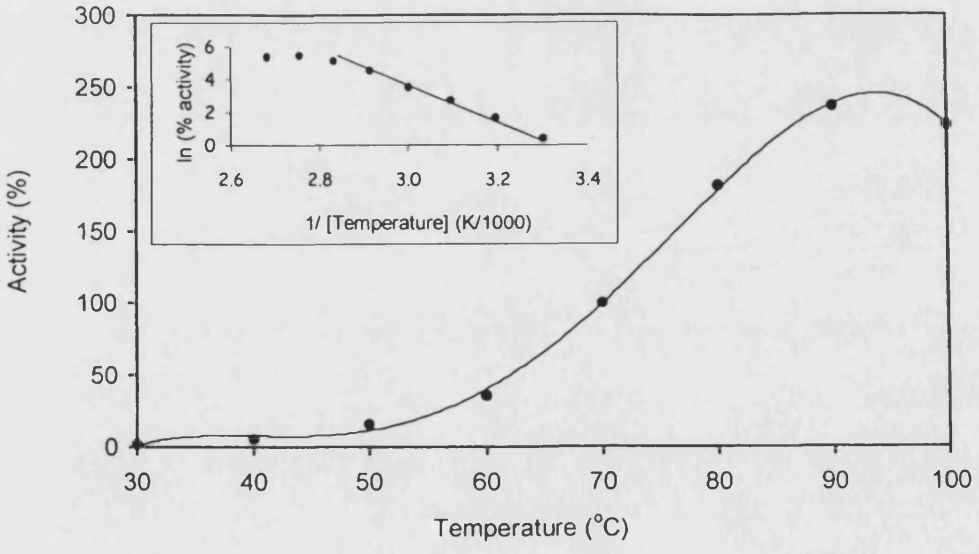


Figure 3 Temperature-dependence of KDG-aldolase activity

Purified recombinant KDG-aldolase was assayed as described in the Experimental section at temperatures from 20 to 100 °C. Activities are expressed as a percentage of the activity at 70 °C, the routine assay temperature. An Arrhenius plot of these data is shown in the inset.

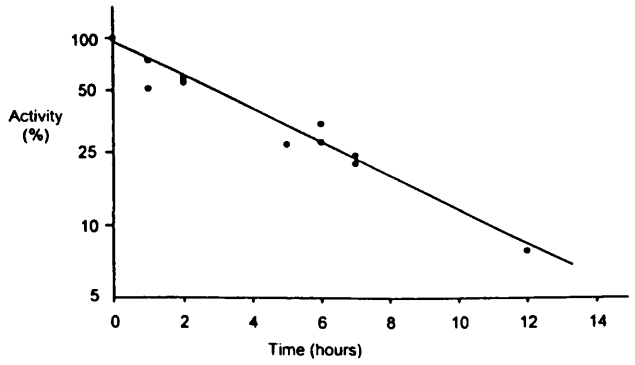


Figure 4 Thermal inactivation of KDG-aldolase from *S. solfataricus*

Purified recombinant KDG-aldolase was incubated at 100 °C for up to 12 h, as described in the Experimental section. After rapid cooling, residual enzymic activity was measured at each time point in the normal aldolase assay.

substrates, the K_m that would be determined from assays with the racemic mixture can be calculated from the rate equation to be:

$$K_m = \frac{2K_m^D \cdot K_m^L}{K_m^D + K_m^L}$$

From the values given in Table 2 for the individual substrates, it was therefore predicted that the value of K_m for the racemic mixture would be 5.0 mM, as indeed was observed.

KDG-aldolase activity was confirmed using the HPLC assay. Incubation of pyruvate and glyceraldehyde in the presence of purified enzyme resulted in utilization of both substrates in equimolar quantities and the concomitant appearance of a product peak with a retention time identical to the KDG marker.

KDG-aldolase is active over a wide range of temperatures (Figure 3), with activity approximately doubling with every 10 °C rise in assay temperature between 30 and 80 °C. An Arrhenius plot of these data is linear over this temperature range (Figure 3, inset), but a break is observed at 80 °C. However, as indicated from the thermostability data described below, the lower-than-expected activity above 80 °C is not due to inactivation of the enzyme (only a 3 % loss of activity is expected at 100 °C over the 10-min period of the assay), and we have shown that there is no significant destruction of the substrates at these temperatures. As expected, therefore, the recombinant enzyme is extremely thermostable under normal assay conditions. This was also found to be the case in the absence of substrates, the enzyme having a half-life of 2.5 h at 100 °C (Figure 4) and 7.8 h at 95 °C, as measured by the rate of irreversible thermal inactivation.

Determination of M_r

Fitting of nine data sets from the equilibrium centrifugation analysis at three different rotor speeds and covering a protein concentration range of 0.1–1.0 mg/ml demonstrated that the recombinant KDG-aldolase exists as a single species of M_r(±S.E.) 133000 (±14000). From a sequence polypeptide M_r of 32980, it is concluded that the enzyme comprises four identical subunits. This was confirmed by sedimentation velocity analysis, which showed a single species of $s^0_{20,w} = 7.0(\pm 0.3)$ S. The diffusion coefficient ($D^0_{20,w}$) of the KDG-aldolase was determined by dynamic laser-light scattering to be 5.4 (± 0.1) $\times 10^{-11}$ m^2/s, giving, via the Svedberg equation, an M_r of 126000 (±6000).

Table 3 Effect of $NaBH_4$ reduction on catalytic activity of KDG-aldolase

KDG-aldolase was incubated with $NaBH_4$ in the presence and absence of the substrates pyruvate and glyceraldehyde, as described in the Experimental section.

Incubation conditions	Activity remaining (%)
Enzyme	100
Enzyme + $NaBH_4$	78
Enzyme + $NaBH_4$ + pyruvate	31
Enzyme + $NaBH_4$ + D,L-glyceraldehyde	90

Reduction with $NaBH_4$

The involvement of a Schiff-base mechanism in the KDG-aldolase reaction was examined by treating the enzyme with sodium borohydride in the presence or absence of substrates pyruvate and D,L-glyceraldehyde in a similar method to that of Aisaka et al. ([15], and references therein). The results shown in Table 3 indicate the participation of a Schiff-base mechanism in catalysis; 69 % inactivation of the enzyme was observed in the presence of $NaBH_4$ and pyruvate, and complete inactivation was observed when the reaction time was increased to 20 min.

Effect of metal ions

Aldolases can be categorized as type-I enzymes, which form a Schiff base with the donor substrate and an active-site lysine residue, and type-II enzymes, which require a metal-ion cofactor. Using the standard assay, KDG-aldolase activity was unaffected by the presence of 0.1 mM $ZnCl_2$ and there was no decrease in activity in the presence of 50 mM EDTA. This is in contrast with the thermostable type-II aldolase from *Methanococcus janaschii*, which was inactivated in the presence of EDTA and re-activated by addition of Zn^{2+} ions [21].

DISCUSSION

Aldolases in biotransformations

In the field of organic synthesis, complex structures are often difficult to prepare by conventional chemical means. Consequently, the use of enzymes to effect specific synthesis under mild reaction conditions, without the need for protecting groups and with chiral control, is gaining in favour. Stereoselective carbon–carbon-bond synthesis is a demanding synthetic challenge in this area, and aldolases, which mediate this reaction through the aldol condensation of an aldehyde acceptor with a ketone donor, are therefore of particular interest [22–24].

Over 30 different aldolases have been identified, and these can be divided into two types. Type-I aldolases, which are found primarily in animals and higher plants, form a Schiff-base intermediate and do not require a metal cofactor, whereas type-II aldolases, found predominantly in micro-organisms, use Zn^{2+} as a cofactor. Fructose-1,6-bisphosphate aldolase, a type-I aldolase from rabbit muscle, has been studied extensively. In common with many other aldolases, mostly type-II enzymes of microbial origin, this enzyme uses dihydroxyacetone phosphate as ketone donor. These aldolases have all been studied to assess their potential as biotransformation catalysts [25].

Clearly, in biotransformation reactions there would be a considerable advantage in using an aldolase that does not require

phosphorylated substrates. Such enzymes are known, one group of which uses pyruvate as the donor substrate. The best characterized is neuraminic acid lyase/aldolase (NAL), a type-I aldolase that catalyses the condensation of pyruvate with *N*-acetyl-D-mannosamine to form *N*-acetylneuraminic acid (sialic acid). Although the enzyme has a catabolic function *in vivo*, synthesis of the aldol product can be achieved by using an excess of pyruvate [26]. NAL is highly specific for pyruvate as donor substrate, but will utilize a wide range of compounds as acceptors, including hexoses, pentoses and tetroses of both D- and L-configuration. In addition to specificity studies [27], the application of NAL in an efficient process for production of neuraminic acid, with potential for scale-up, has been described [28].

In addition to its substrate specificity, a second important feature for any enzyme used in a biotechnological process is its stability, which not only limits the conditions under which it can be used but may determine to a large extent the economics of the particular process. In this regard, extremozymes, which are enzymes from organisms growing in extreme environments (extremophiles), may be particularly useful (reviewed in [29]). That is, extremozymes possess stabilities commensurate with the growth conditions of the parent organism, and extremophiles are therefore a rich source of novel and hyperstable enzymes.

KDG-aldolase from *S. solfataricus*

In the context of these requirements, we have chosen to characterize the enzyme KDG-aldolase from the hyperthermophilic Archaeon, *S. solfataricus*. This aldolase uses non-phosphorylated substrates, pyruvate and glyceraldehyde, and was predicted to be highly thermostable, since *S. solfataricus* grows optimally at 80–85 °C. Moreover, the product of the reaction, 2-keto-3-deoxygluconate, contains four different oxidation states of carbon in four contiguous carbons, and therefore provides a framework for a number of sugar transformations in aqueous solution without the need for protecting groups [22].

To this end, we purified the *S. solfataricus* KDG-aldolase and, from the determined N-terminal sequence, were able to generate by PCR amplification an oligonucleotide probe that then allowed us to clone and sequence the gene. Expression in *E. coli* permitted the easy purification of active recombinant enzyme that was then shown to be a tetrameric protein with a high thermal stability.

Kinetically, the enzyme was indistinguishable from the native KDG-aldolase purified from *S. solfataricus* cells. One interesting feature with respect to the enzyme's thermostability and thermoactivity is that the enzyme shows the expected increase in catalytic activity with temperature up to 80 °C, the growth temperature of the parent organism, but then the activity is considerably less than expected at 90 and 100 °C. Over the time period of the assay, there is no significant thermal inactivation of the enzyme at these temperatures, and therefore it is possible that the enzyme undergoes a reversible dissociation or unfolding affecting catalytic activity. This effect has also been proposed for *Thermoanaerobacter* 3-phosphoglycerate kinase [30], and emphasizes the point that thermostability alone does not guarantee thermoactivity.

Relationship to other aldolases

Analysis of the translated amino acid sequence has shown the *S. solfataricus* KDG-aldolase to be a putative member of the NAL enzyme superfamily of Schiff-base-dependent aldolases, dehydratases and decarboxylases [18], showing 28 and 29% identity respectively with NAL and DHDPS. Both NAL and DHDPS catalyse reactions involving C–C bond formation with pyruvate as the ketone donor, utilizing a protonated lysine to form a Schiff-base intermediate with the keto-group of the substrate. This lysine is conserved in all members of the superfamily and has been identified in the three-dimensional structure of both NAL and DHDPS. A lysine in the sequence of KDG-aldolase (K155; Figure 2) aligns well with this conserved lysine residue, suggesting that the *Sulfolobus* enzyme employs a similar reaction mechanism. As in the case of the other type-I aldolases, this has been confirmed by the sodium borohydride-induced inactivation of the enzyme in the presence of pyruvate. Furthermore, KDG-aldolase activity is not affected by the presence of EDTA or Zn^{2+}, confirming that it is not a member of the type-II (metal-ion-dependent) family. Finally, the enzyme was shown to be a tetramer, a further characteristic of the type-I aldolases that is distinct from the type-II enzymes, which are generally dimeric.

The *S. solfataricus* KDG-aldolase shows no significant sequence similarity to microbial 2-keto-3-deoxyphosphogluconate aldolases, which catalyse the aldol condensation of pyruvate and glyceraldehyde 3-phosphate, and the enzyme shows no activity with glyceraldehyde 3-phosphate as substrate. The three-dimensional structure of 2-keto-3-deoxyphosphogluconate aldolase from *Pseudomonas putida* has been determined [31]. It also has an α/β barrel structure, but in this case the enzyme is a trimer. The reaction mechanism again proceeds via a Schiff-base intermediate, involving an active-site lysine residue, and appears to resemble that of mammalian fructose-bisphosphate aldolase.

Gene sequence

The protein-coding sequence of the KDG-aldolase gene was identified by comparison of the translated base sequence with the determined N-terminal amino acid sequence, where an exact match was found. However, we could not find a recognizable Shine–Dalgarno sequence upstream of the start codon, nor a possible promoter site in the −25 to −50-bp region. Moreover, there is no good potential transcriptional stop signal (poly-dT tract) downstream of the KDG-aldolase open reading frame (ORF). The absence of these regions is suggestive of the aldolase gene being part of an operon structure, and translation of the flanking sequences is consistent with this possibility. The upstream sequence (97 bp) translates into a potential ORF that terminates 2 bp before the aldolase ATG start codon; however, no matches to this partial sequence were found in the databases. A second potential ORF (180 bp) begins 9 bp downstream of the aldolase gene, and this partial sequence, when aligned with the protein databases, shows the highest identity score with a *Bacillus* fructokinase (41%). The metabolic significance of the flanking genes and the possible operon structure will only become apparent when further and more extensive sequencing is carried out; clearly, this is beyond the scope of the current paper.

Concluding remarks

The significance of the *S. solfataricus* KDG-aldolase in biotransformations and organic syntheses lies in its thermostability and its catalytic activity with non-phosphorylated substrates. Future applications may require the manipulation of the enzyme structure to recognize other substrates, and to this end we have crystallized the KDG-aldolase and collected diffraction data to 2.2 Å (E. Hendry, C. L. Buchanan, D. W. Hough, M. J. Danson and G. L. Taylor, unpublished work). Structural analysis is in progress.

We thank Professor R. Eisenthal (University of Bath, Bath, U.K.) for guidance with the kinetic analyses. We also thank Professor R. Sharp and Dr. N. Raven (Centre for Applied Microbiology and Research, Porton Down, Wilts, U.K.) for providing the *S. solfataricus* cell pastes, Mrs. J. Young (Zeneca Pharmaceuticals, Macclesfield, U.K.) for amino acid sequencing, Dr. L. Irons (University of Bath) for help with the determination of the diffusion coefficient, Mr. C. Badcock (University of Bath) for help with the thermostability experiments, and Dr. A. Rowe and Dr. N. Errington (U.K. National Centre for Macromolecular Hydrodynamics, University of Nottingham, Nottingham, U.K.) for the hydrodynamic analyses. This work was funded by a Biotechnology and Biological Sciences Research Council (U.K.) studentship to C. L. B.

REFERENCES

1 Grogan, D. W. (1989) J. Bacteriol. **171**, 6710–6719
2 De Rosa, M., Gambacorta, A., Nicolaus, B., Giardina, P., Poerio, E. and Buonocore, V. (1984) Biochem. J. **224**, 407–414
3 Budgen, N. and Danson, M. J. (1986) FEBS Lett. **196**, 207–210
4 Siebers, B. and Hensel, R. (1993) FEMS Microbiol. Lett. **111**, 1–8
5 Selig, M., Xavier, K. B., Santos, H. and Schonheit, P. (1997) Arch. Microbiol. **167**, 217–232
6 Laemmli, U.K. (1970) Nature (London) **227**, 680–685
7 Skoza, L. and Mohos, S. (1976) Biochem. J. **159**, 457–462
8 Gottschalk, G. and Bender, R. (1982) Methods Enzymol. **90**, 283–287
9 Bradford, M. M. (1976) Anal. Biochem. **72**, 284–255
10 Connaris, H., West, S. M., Hough, D. W. and Danson, M. J. (1998) Extremophiles **2**, 61–66
11 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Press, Cold Spring Harbor
12 Sanger, F., Nicklem, S. and Coulson, A. R. (1977) Proc. Natl Acad. Sci. U.S.A. **74**, 5463–5467
13 Johnson, M. L., Correia, J. J., Yphantis, D. A. and Halvorson, H. R. (1981) Biophys. J. **36**, 575–588
14 Claes, P., Dunford, M., Kenney, A. and Vardy, P. (1992) In Laser Light Scattering in Biochemistry (Harding, S.E., Sattelle, D. B. and Bloomfield, V. A., eds.), pp. 66–76, Royal Society of Chemistry, Cambridge
15 Aisaka, K., Igarashi, A., Yamaguchi, K. and Uwajima, T. (1991) Biochem. J. **276**, 541–546
16 Atschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) J. Mol. Biol. **215**, 403–410
17 Izard, T., Lawrence, M. C., Maltby, R. L., Lilley, G. G. and Coleman, P. M. (1994) Structure **2**, 361–369
18 Babbitt, P. C. and Gerlt, J. A. (1997) J. Biol. Chem. **272**, 30591–30594
19 Lawrence, M. C., Barbosa, J. A. R. G., Smith, B. J., Hall, N. E., Pilling, P. A., Ooi, H. C. and Marcuccio, S. M. (1997) J. Mol. Biol. **226**, 381–399
20 Wharton, C. W. and Eisenthal R. (1981) Molecular Enzymology, Blackie and Son, Glasgow
21 Choi, I.-G., Cho, S. K. and Yu, Y. G. (1998) J. Biochem. Mol. Biol. **31**, 130–134
22 Shelton, M. C., Cotterill, I. C., Novak, S. T. A., Poonwala, R. M., Sudarshan, S. and Toone, E. J. (1996) J. Am. Chem. Soc. **118**, 2117–2125
23 Takayama, S., McGarvey, G. J. and Wong, C.-H. (1997) Ann. Rev. Microbiol. **51**, 285–310
24 Fessner, W.-D. (1998) Curr. Opin. Chem. Biol. **2**, 85–97
25 Wong, C.-H., Halcomb, R. L., Ichikawa, Y. and Kajimoto, T. (1995) Angew. Chem. Int. Ed. Engl. **34**, 412–432
26 Uchida, Y., Tsukada, Y. and Sugimori, T. (1984) J. Biochem. (Tokyo) **96**, 507–522
27 Fitz, W., Schwark, J.-R. and Wong, C.-H. (1995) J. Org. Chem. **60**, 3663–3670
28 Mahmoudian, M., Noble, D., Drake, C. S., Middleton, R. F., Montgomery, D. S., Piercey, J. E., Ramlakhan, D., Todd, M. and Dawson, M. J. (1997) Enz. Microb. Technol. **20**, 393–400
29 Hough, D. W. and Danson, M. J. (1999) Curr. Opin. Chem. Biol. **3**, 39–46
30 Thomas, T. M. and Scopes, R. K. (1998) Biochem. J. **330**, 1087–1095
31 Mavridis, I. M., Hatada, M. H., Tulinsky, A. and Lebioda, L. (1982) J. Mol. Biol. **162**, 419–444
32 Higgins, D. G. and Sharp, P. M. (1988) Gene **73**, 237–244
33 Eisenthal, R. and Cornish-Bowden, A. (1974) Biochem. J. **139**, 715–720

Received 11 March 1999/29 June 1999; accepted 3 August 1999